CUK-H05989-19-KP2664
04/01/06

THE JOURNAL OF PHILOSOPHY

⑲

VOLUME CII, NUMBER 7
JULY 2005

page

327 *Evaluating Social Reasons: Hobbes versus Hegel* — Anthony Simon Laden

357 *Is There a Single Objective, Evolutionary Tree of Life?* — Joseph LaPorte

COMMENTS AND CRITICISM

375 *A Note on Goodman's Problem* — Robert Schwartz

380 NEW BOOKS: TRANSLATIONS

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Christopher Peacocke, Philip Pettit, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.

You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2005

SUBSCRIPTIONS (12 issues)

KP 2564

Individuals	$45.00
Libraries and Institutions	$100.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.

POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2005 by the Journal of Philosophy, Inc. ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY
VOLUME CII, NO. 7, JULY 2005

EVALUATING SOCIAL REASONS: HOBBES VERSUS HEGEL*

When we deliberate together, we offer reasons to one another and evaluate those reasons in the process of accepting or rejecting them. Sometimes we appeal to the norms or characteristics of groups of which we are both members, or to features of our particular relationship. We argue against a proposed law by explaining why it would be unconstitutional. We defend our actions against charges that they are unfair or exploitative. In such cases, we invoke and evaluate and respond to what I will call *social reasons.*[1] Social reasons derive their authority from the features of the norms and rules that govern and constitute various groups, institutions, and

---

* This paper was originally written for a conference on "Reason and Deliberation" held at Bowling Green State University in 2001 under the title "Reasoning Together à la Hobbes and à la Hegel." I gave a shorter version with that title at a one-day conference of the Chicago Political Theory Group in November 2002. I benefited from the discussion on both occasions. I would also like to thank an anonymous referee for helpful comments on an earlier draft.

[1] The importance of such reasons has been stressed by a number of philosophers in recent years. See, for instance, Samuel Freeman, "Reason and Agreement in Social Contract Views," *Philosophy and Public Affairs*, XIX, 2 (Spring 1990): 122–57, who calls such reasons "public," and uses them to distinguish Hobbesian from Rousseauvian social contract theories; Christine Korsgaard, *The Sources of Normativity* (New York: Cambridge, 1996), and "The Reasons We Can Share," reprinted in *Creating the Kingdom of Ends* (New York: Cambridge, 1996), pp. 275–310, where she refers to reasons as intersubjective and finds their original articulation in Kant; Charles Taylor, "Explanation and Practical Reason," "Irreducibly Social Goods," and "To Follow a Rule," all reprinted in *Philosophical Arguments* (Cambridge: Harvard, 1995), pp. 34–60, 127–45, 165–80, where he talks of "common understandings" and traces their importance in practical reasoning to Hegel and Wittgenstein; and Onora O'Neill, "Four Models of Practical Reasoning," reprinted in *Bounds of Justice* (New York: Cambridge, 2000), pp. 11–28, where she contrasts "action-based" accounts of practical reason with "teleological" ones, tracing the former back to Hegel, Wittgenstein, and Kant, and the latter to Plato and Hume. I have described such reasons elsewhere as "we"-reasons. See my "Outline of a Theory of Reasonable Deliberation," *Canadian Journal of Philosophy*, XXX (December 2000): 551–80.

0022-362X/05/0207/327–56

relationships, which I will corral under the heading of "plural subject."[2] The evaluation of such reasons is both practically and theoretically complicated. At one level, such reasons must find support in the norms to which they appeal. Such norms are not necessarily fixed, however, nor is our allegiance to them necessarily permanent or obligatory. Thus, we must sometimes face the further questions of whether those norms are justified, on what basis they can be justified, and whether that justification for us can survive challenges to the reason in question that come from the norms of other plural subjects to which we also belong. Moreover, if some of the reasons we face arise from different sources, evaluating those reasons will require overcoming the at least prima facie incommensurability between competing reasons.

In this paper, I contrast a broadly Hobbesian approach to the evaluation of social reasons with a broadly Hegelian approach, and argue for the merits of the Hegelian approach. Though I call the second approach Hegelian, others advocating similar approaches recently have described their views as Kantian or Rousseauvian.[3]

The Hegelian approach differs from the Hobbesian approach along two different axes, one structural and the other substantive. The structural axis determines the appropriate standpoint from which to evaluate social reasons. Along this axis, the move from the Hobbesian to the Hegelian approach is both small and natural, and in important ways mirrors the move from orthodox rational choice theory to the Hobbesian approach. Thus, the comparison of the two approaches along the structural axis provides an argument for the Hobbesian to adopt the Hegelian position. What, I think, prevents many a Hobbesian from taking this step, however, is the difference between the two theories along the substantive axis. The substantive axis concerns the basic ingredients of a practical reason. Here, the difference between the Hegelian and Hobbesian approaches is more pronounced. After laying out this difference, I argue, first, that the Hobbesian positions along the two axes are related and that the substantive position helps to justify the structural position, and second, that some of the more common objections to the Hegelian's substantive position are either mistaken or not as forceful as they first appear.

---

[2] I borrow the term "plural subject" from Margaret Gilbert, who uses it in her analysis of collective action and intention. See, for instance, her *On Social Facts* (Princeton: University Press, 1989). I use it in a somewhat looser fashion, though in the same spirit: roughly, a plural subject exists any time it is appropriate for its members to refer to it using the first person plural.

[3] See, for instance, Freeman, "Reason and Agreement in Social Contract Views"; Taylor, "Explanation and Practical Reasons," and "Irreducibly Social Goods"; O'Neill, "Four Models of Practical Reason"; and Axel Honneth, *The Struggle for Recognition*, Joel Anderson, trans. (Cambridge: MIT, 1996). Note also that while the Hegelian approach draws on some resources to be found in the work of Hegel, it does not exhaust them. My reason for drawing on some but not all of the resources in Hegel is that those I draw on are less intricately connected to Hegel's full metaphysical system and thus are more accessible to those of a Hobbesian persuasion.

The aim of this paper is neither the ambitious one of mounting an overall defense of a Hegelian approach to practical reasoning nor the preliminary one of showing that the category of social reasons is not empty. Rather, I aim to show that if we accept the existence and force of social reasons as even the Hobbesian does, then a Hegelian approach provides us with a better answer to the very practical questions of how to evaluate the strength of social reasons in the course of reasoning together. Doing so helps to clarify the nature of collective deliberation, and the way that such deliberation both tracks and generates reasons. It can thus provide some insight into when the urging of certain claims or their rejection is reasonable, where this question is not reducible to the question of which actions, whether individual or collective, are rational.[4]

Because this paper deals with the difference between basic approaches rather than particular theories, it of necessity moves at a high level of generality, and this can lead to misunderstanding. In particular, there are many different ways of proceeding along either of the two approaches I highlight and my argument will, in general, not pay attention to these basic differences. Nevertheless, by way of illustration, I will from time to time explain a point by reference to a particular theory that adopts a given approach. In those cases, a proponent of a different theory within the same approach is likely to reply that a more sophisticated version of that approach will not run into that particular danger. I ask such a reader for a grain or two of salt at those moments, to see the smaller argument as an illustration of the larger point, not the full argument for it, and to ask whether the more general point being made nevertheless applies to her favored position.

Before turning to the main arguments, it will help to have some examples before us. Consider three:

George and Alice are friends. Even though George's reasons for hanging around with Alice originally had to do with the fun they had together, he has, over time, developed a whole new set of reasons for spending time with her. Their friendship gives him reasons to spend time with her that are not always subject to considerations about his own interests. When Alice's life goes through a difficult period, and she calls on George to be there for her, she need merely say to him, "George, I need a friend." If George is really her friend, then no other reason need be given. George, however, is not Alice's slave. He can acknowledge that his friendship with Alice supports this kind of claim while rejecting her particular claim here and now. Perhaps her need for him at the moment conflicts with other demands on his time, demands that come from personal projects or other relationships. How, then, is George to evaluate Alice's claim of friendship in a particular instance, and when is his rejection of those demands reasonable?

---

[4] It thus also has affinities with the contractualist view of T.M. Scanlon. See his *What We Owe to Each Other* (Cambridge: Harvard, 1999). In particular, I think the Hegelian approach discussed below provides a helpful framework for thinking about the appropriate criteria for reasonable rejection within Scanlon's theory. On the difference between theories of rational choice and theories of reasonable deliberation, see my "Outline of a Theory of Reasonable Deliberation."

The second example involves what John Rawls calls public reasons.[5] Consider several ways I might make a case in favor of some or other principle of distributive justice. I could argue that my favored principle would advance each of my fellow citizens' original, individual interests. Alternatively, I could appeal to our common citizenship and argue that one of the conditions of our continuing to be citizens together is some level of material equality. Again, as with the appeal to friendship, this argument rests on a social reason, and so we can ask how my fellow citizens should evaluate its force, and what sorts of considerations they can reasonably bring to bear in either endorsing or rejecting it.

Finally, consider representatives of the member states of the European Union when they deliberate together about the adoption of a particular regulatory scheme. In the course of that deliberation, some reasons will be directed at the representatives of particular countries and point out how the proposal is consistent with or even advances their national interest. Others might call on member states to honor their explicit commitments made in various European treaties. But some of the reasons offered in the deliberation will appeal to the shared Europeanness of the representatives. They will be asked to act not on the basis of national interest but for something akin to the common European good. An appeal will be made to their identity as Europeans and not merely representatives of particular member countries, and the question will once again arise: How should they evaluate the force of these reasons?

---

[5] See, for instance, Rawls, "The Idea of Public Reason Revisited," reprinted in *Collected Papers*, Freeman, ed. (Cambridge: Harvard, 1999), pp. 573–615.

This example brings out two further features of social reasons that are not obvious in my first two examples, and which further complicate the practical question of evaluation. First, while these reasons appeal to the deliberators' collective membership in a plural subject, they call on each deliberator to place the demands of that membership over those placed on her by other plural subjects of which she is a part. The French MEP in the above example is being asked to put her European identity above her French identity in this scenario. Her individual interests are not so obviously on the table. Second, sometimes the offering of a particular social reason serves more as an invitation than a call to duty. Citizens of the member states of the EU do not yet form a robust plural subject as Europeans, and so when reasons are offered in European contexts that appeal to this shared identity, they also participate in the process of bringing into being such a plural subject.

I. STRUCTURAL DIFFERENCES

The first axis along which the two approaches differ is what they take to be the appropriate standpoint for evaluating social reasons. An individual agent adopts an individual evaluative standpoint when she asks whether a given course of action would be good for her. She adopts a particular social standpoint when she asks if it would be good for a particular plural subject of which she is a member. We take up a social evaluative standpoint when we engage in what Robert Sugden and Michael Bacharach have called team reasoning.[6] Different theories of practical reason within each approach will, of course, provide different accounts of what makes it the case that a course of action is "good for" either an agent or a plural subject. What, I argue, distinguishes the Hegelian from the Hobbesian approach is the role it assigns the individual evaluative standpoint. Whereas the Hobbesian approach takes such a standpoint to be fundamental to the evaluation of all social reasons, the Hegelian approach does not. Despite that characterization of the difference between the two approaches, I hope to show that once one has adopted the Hobbesian approach, there is no reason, on structural grounds, not to adopt the Hegelian approach.

---

[6] See Sugden, "Team Preferences," *Economics and Philosophy*, XVI (2000): 175–204; and Bacharach, "Interactive Team Reasoning: A Contribution to the Theory of Co-operation," *Research in Economics*, LII (1999): 117–47. In team reasoning, each member of a team figures out what to do by figuring out what set of moves would jointly produce the best result for the team, and then doing her part.

The argument recalls the strategy of Thomas Nagel's argument in the *Possibility of Altruism.*[7] Nagel argues that the desires of others give us reasons to act by showing, first, that our own desires at times other than the present give us reasons to act. He then maintains that the same sorts of considerations that push us from a present-desire theory of reasons to a requirement of prudence push us from a self-interested theory to an altruistic one. Accepting prudence but not altruism, Nagel maintains, is a kind of halfway position that argumentative momentum renders unstable. Similarly, along the structural axis, the considerations that lead the Hobbesian to reject orthodox rational choice theory turn out to push past the Hobbesian approach to the Hegelian one. Orthodox rational choice theory denies that there are any truly social reasons, claiming that all reasons must be evaluated directly from an individual standpoint. The Hobbesian approach rejects this claim. It accepts the existence of social reasons that have a certain independent force, and claims that I should evaluate these reasons, first, from the standpoint of the plural subject whose reason they purport to be, and derivatively, in terms of the reasons I have for being a member of that plural subject. Nevertheless, it further claims that the only appropriate standpoint for evaluating reasons for being a member of a given plural subject is an individual one. The Hobbesian thus admits of multiple evaluative standpoints, but insists that the individual standpoint occupies a privileged, foundational role. The Hegelian approach denies that the individual standpoint provides the only valid ultimate grounds for membership in a plural subject. The claim, then, is that once we accept the independent force of social reasons, we should also take them as a potential source of reasons for our membership in plural subjects, and thus a potential source for our evaluation of other social reasons. That is, the considerations that push us to take the step from orthodox rational choice theory to the Hobbesian position also push us further along that path to the Hegelian position.[8]

---

[7] Princeton: University Press, 1970.

[8] Note that although the strategy is the same, the argument pursued in this section is somewhat orthogonal to Nagel's. He aims to show that once you move from a theory which grounds reasons only in present desires to one which grounds reasons in desires at any time, it is hard to resist the further step that acknowledges the possibility that the desires of others also ground reasons for us. I aim to show that once you accept that social reasons have some independent force, it is hard to resist the further step that accepts that we can evaluate some social reasons in terms of other ones. I assume throughout what Nagel calls the "temporal neutrality" of individual reasons, the very point Nagel insists requires argument. Although I say nothing in this paper about what the Hegelian approach might have to say about the move from a present-desire theory to a prudential theory, I do think the approach discussed here provides resources for thinking about that question as well. See my "Prudence Reconceived" (unpublished manuscript).

I.A. From Rational Choice Theory to Hobbes. We might, echoing Bernard Williams, describe orthodox rational choice and decision theory as a sub-Hobbesian view. According to this view, the only appropriate standpoint from which to evaluate reasons offered to an agent is that agent's individual standpoint. Depending on the version of the sub-Hobbesian approach one takes, such evaluation might proceed in terms of the agent's preferences or utilities or interests. In each case, however, the agent asks "Is it good for me?" in evaluating a claim on her. In essence, this means that the sub-Hobbesian approach denies that there are any truly social reasons. Social reasons will only count as reasons on this approach if they are merely individual reasons in disguise. Of course, a sub-Hobbesian position can be sophisticated, and accept that, for example, an individual's utility is affected by what happens to at least certain others, perhaps via something like sympathy. What is important for my purposes here, however, is that even the sympathetic sub-Hobbesian agent only ever asks, when evaluating a reason, is it good for me that I accept this claim? The sub-Hobbesian will thus say that George, faced with Alice's claims on him, should evaluate them from his individual standpoint. It might turn out that George's sympathy for Alice makes it the case that he is better off sacrificing some of his other projects to lessen her suffering—he need not be hard-hearted. But in the absence of such factors that inform George's individual standpoint, the sub-Hobbesian view will conclude that Alice's need of a friend is not a reason for George at all.[9]

Sub-Hobbesian individualism about reasons does not, however, generate an interesting approach to reasoning together. To some degree, it denies the very possibility of really reasoning together. If we insist on always evaluating the claims of others from our individual standpoint, we treat others not as deliberative partners but as complicated parts of our choice environment. We can reason alongside such entities, but we cannot reason with them. Moreover, as Hobbes observed, if we recognize no social reasons with independent normative force, then we have no way to prevent the war of all against all. What, then, separates out Hobbes and especially a neo-Hobbesian like David Gauthier from more orthodox rational choice theorists is that they support some appeals to a social evaluative standpoint.[10]

---

[9] There is another question which lies in these waters and from which it is important to distinguish the question that concerns me here: To what degree must the interests that ultimately ground reasons be subjective? According to some formulations of rational choice theory, reasons must find support in the satisfaction of preferences or a change in utility. Such theories will have to account for George's concern for Alice with something akin to sympathy. Amartya Sen accuses such theories of describing "rational fools" and suggests that an adequate account of rationality will have to take account of more objective sources of reasons, such as what he calls commitments. See his "Rational Fools," reprinted in *Choice, Welfare, and Measurement* (Cambridge: MIT, 1982), chapter 4. Sen's concern is with how broadly we conceive of individual interests, and thus with what can inform the individual evaluative standpoint. Mine, in contrast, is with how to evaluate reasons that do not appeal to that standpoint, however broadly construed. Clearly there is some overlap between the questions, as the more broadly we construe the individual standpoint, the wider the terrain of reasons it can support. Nevertheless, the sub-Hobbesian could accept objective sources of individual interest without thereby abandoning his claim that all reasons must be evaluated from the individual standpoint.

Both Hobbes and Gauthier start from a basic insight: it can be rational for us to keep our agreements even when doing so will constrain the pursuit of our short-term interests. The thought behind this claim is that by being the sort of people who can be trusted to keep agreements, we will open up for ourselves a vast array of cooperative opportunities whose advantages in the long term will outweigh the short-term advantages from breaking agreements. In order to be the kind of person who can keep such agreements, however, I have to be willing to adopt a social evaluative perspective when such reasons are on the table. If, to take an example from Rousseau, we have agreed to work together to catch a deer in order to feed ourselves, and an opportunity arises for me to feed myself by abandoning our joint project, I have to be able and willing to evaluate that opportunity from the standpoint of our joint endeavor, and not merely from my individual point of view. So the Hobbesian admits that there are times when it is appropriate to evaluate social reasons from a social standpoint. At the same time, however, the Hobbesian's grounds for accepting such evaluation is, ultimately, based on an appeal to the individual standpoint. I should act on social reasons when acting on that kind of social reason is, all-tolled, good for me. Thus, standing behind my adoption of social evaluative standpoints is always my individual standpoint. And this provides the Hobbesian with the following general rule for evaluating social reasons: take up a social standpoint for the evaluation of social reasons only insofar as being a member of the plural subject that could generate such reasons is to my advantage. It is in this sense that the Hobbesian gives the individual standpoint a privileged, foundational role in the evaluation of social reasons.

---

[10] Hobbes, *Leviathan*, Edwin Curley, ed. (Indianapolis: Hackett, 1994); Gauthier, *Morals by Agreement* (New York: Oxford, 1986). There is a deep question of Hobbes interpretation here that I want to bracket. Hobbes's belief in the kind of indirect reduction that lies at the heart of Gauthier's work is displayed most prominently in the argument against the Fool in *Leviathan*. His more orthodox tendencies appear most clearly in his claims that the laws of nature do not bind in the absence of an enforcer. In calling the approach outlined in this section Hobbesian, I mean to invoke the Hobbes who argues against the Fool. The more orthodox game-theoretic work of someone like Ken Binmore also claims to be Hobbesian, but allies itself with the other aspect of Hobbes's view. See his *Game Theory and the Social Contract, Volume I: Playing Fair* and *Volume II: Just Playing* (Cambridge: MIT, 1994).

Such an approach is not without explanatory power, as Gauthier and others have amply demonstrated. Even if the force of social reasons must ultimately be evaluated from an individual standpoint, it is possible that my membership in a given plural subject can generate types of reasons for me that I would not otherwise have had. In addition, it may be the case that, taken on their own, some of these reasons would not meet with my endorsement from my individual standpoint. So, for instance, I can only be a friend if, in the evaluation of my friend's claims on me, I rely on the norms of friendship and not merely my individual interests. Since my life will go better with friends than without them, I can, from my individual standpoint, endorse taking up the standpoint of friend when evaluating the claims of particular friendships. Thus, on this account, we can explain why at least some social reasons have force even when they themselves do not appeal to our self-interest.[11]

Go back to my first example. Whether or not George is better off spending the evening listening to Alice, it is certainly likely that he is better off having a friend like Alice in his life. So his evaluation of her claim has, at least implicitly, two stages. At the first stage, he asks whether the norms of their friendship support her claim on him. Here, he adopts a social evaluative standpoint, that of their friendship. At the second stage, he asks whether it is good for him to have friendships like that. Here, he adopts his individual evaluative standpoint. If George opts for the friend-filled life, he must also recognize the force of the social reasons that arise out of his friendships. These, then, are a new kind of reason. Their tie to his individual interests is indirect, and they cannot be understood or properly evaluated without recognition of this fact. Note here that what is new is the *kind* of reason friendship offers. According to some sub-Hobbesian approaches, friendships can generate new reasons insofar as they, for example, develop the effects of sympathy. In that case, however, the reasons our friends generate are not new *kinds* of reasons. Their force does not arise from the adoption of a new evaluative standpoint.

---

[11] It will help some readers at this point to draw a connection between this approach and Rawls's discussion of practices in "Two Concepts of Rules," *The Philosophical Review*, LXIV, 1 (1955): 3–32, reprinted in *Collected Papers*, pp. 20–46. Social reasons are reasons internal to a practice. Their immediate authority stems from the constitutive features of the practice. The Hobbesian claims, however, that only practices that I can support from my individual evaluative standpoint can generate reasons with normative force. Thus, while the reasons within a practice need not appeal to our individual interests, the reasons that support the practice as a whole do.

At the same time, there are some sorts of friend-like relationships we are better off without. There may come a point when Alice's demands become exploitative and George concludes that he'd rather have no friends than friends like that. The Hobbesian approach gives us a clear-cut method for identifying that point. Not all relationships can be endorsed from my individual standpoint. Whether or not a relationship can be so endorsed may depend on what sorts of reasons and demands it authorizes. Hobbes and Locke, for instance, argue that we can determine the limits of state power by asking whether a state with a particular power does in fact advance each individual's interests in comparison with life in the absence of a state.[12] Similarly, the Hobbesian friend asks whether this friendship, given its particular contours and demands, is good for him. Just as we can reject certain demands of certain states without thereby rejecting the possibility of any social organization, we can reject the particular demands of some friends without ruling out the possibility of any true friendship.[13]

The Hobbesian approach to reasoning together relies on the existence of a multiplicity of evaluative standpoints. Within each particular relationship, group, or practice, the good of the plural subject in question provides an evaluative standpoint from which we can judge the force of particular social reasons. Nevertheless, each of these standpoints is ultimately dependent on a more fundamental standpoint: that of the individual. It is only from this standpoint that my various attachments to plural subjects can themselves be evaluated. Thus, when we are faced with conflicting reasons that stem from our attachment to distinct plural subjects, the Hobbesian approach gives us a method for addressing the conflict: evaluate those attachments from my individual standpoint.

I.B. From Hobbes to Hegel. The move from the sub-Hobbesian to the Hobbesian approach involves challenging the assumption that the individual standpoint is the only appropriate point from which to evaluate reasons, even those internal to a given plural subject. The move from the Hobbesian to the Hegelian approach involves a single further step: raising a similar question about the standpoint from which we evaluate our attachment to plural subjects. The Hobbesian claims that we only have reason to belong to plural subjects that we can endorse from our individual standpoint. The Hegelian, on the other hand, says that we can evaluate our attachment to plural subjects from individual or social standpoints. That is, not all evaluation of social reasons has to trace back to evaluation from the individual standpoint. Just as the evaluation of reasons within a plural subject can appeal to a social standpoint, the evaluation of my attachment to that plural subject can also appeal to a social standpoint.

---

[12] Hobbes, *Leviathan*, p. 110 (chapter 18, pars. 1–2); John Locke, *Second Treatise of Government*, C.B. MacPherson, ed. (Indianapolis: Hackett, 1980), p. 111 (§222).

[13] The Hobbesian need not assume that it is up to each individual to shape her relationships in any way that suits her. There may be a relatively fixed and limited number of ways of being a friend, and what constitutes being a friend will be dependent on existing social norms, not on the agent's particular preferences and needs.

The Hegelian approach relies on two related shifts in the conceptualization of the nature of the relationship of an individual agent to the plural subjects to which she belongs. First, as Hegel puts the point in *The Philosophy of Right*, a distinctive feature of the institutions of *Sittlichkeit* (the family, civil society and the state) is that people belong to these institutions not as individuals but as members.[14] That is, our participation in these institutions changes who we are rather than merely giving us a new list of rights and responsibilities. In contemporary jargon, they give us new practical identities: new senses of who we are, how we value ourselves, and what makes our actions worth undertaking.[15]

So, for example, to be a member of a family is not only to have a set of obligations and privileges with regard to certain other people. It is to conceive of yourself as in part constituted by the role you play in the family, as a wife or father or child, and to regard some of your reasons for acting as arising out of that role. Being a father means that I have reasons for acting one way rather than another that appeal to that role, and that I find certain actions worth undertaking because they are what fathers do, or what I think, as a father, they ought to do. It further means that I do not regard these considerations as outside impositions on my range of choices, but as expressive of who I am.[16] Evaluating reasons grounded in such roles is thus not a matter of reducing social reasons to a single evaluative standpoint provided by individual interests, but involves, in part, interpreting our various memberships and asking substantive questions about them. That we inhabit these institutions as members changes the nature of the evaluation of those roles. While I can still ask, "Why be a father?" and "Should fatherhood place these claims on me?" it is a mistake to understand those questions as calling for me to take up my individual evaluative standpoint. Part, perhaps much, of the value of being a father can only be grasped as of value by someone who is already in that role. From an individual standpoint, the value of fatherhood is likely to be missed or distorted or obscured. In making this claim, I do not mean to deny that part of what makes fatherhood a role I can affirmatively occupy is that it is, in some sense, good for me. Rather, the point is that the "me" for whom it is good cannot be properly characterized without reference to my being a father. And so when we try to evaluate reasons that stem from our membership in plural subjects, there is no reason to privilege the answers arrived at from an individual standpoint that stands apart, as it were, from all our memberships in plural subjects.

---

[14] Hegel, *Elements of the Philosophy of Right*, H.B. Nisbet, trans., Allen Wood, ed. (New York: Cambridge, 1991), §158.

[15] The term "practical identity" comes from Korsgaard, *The Sources of Normativity*. She describes practical identity in basically the terms above on p. 101. I will have more to say about the role played by practical identity in a Hegelian theory of evaluating social reasons together at the end of the paper.

[16] This, of course, raises the question of what to say about cases where a social reason is presented to me on grounds that I belong to a plural subject, where I would reject such membership, or where despite the accuracy of the ascription, it is not one I happily assent to. Clearly, this possibility and its apparent structural similarity to the kinds of cases I have been discussing raises serious problems for any broadly Hegelian account. Nevertheless, I bracket them here, as I think a full description of the Hegelian picture even in the ideal situation where all membership in plural subjects are endorsed and affirmed, is sufficiently difficult and controversial

The second shift involves changing our focus away from the relation between an agent's individual standpoint and her membership in a single plural subject towards the multiplicity of plural subjects of which any one agent is a member. Recognizing the multifaceted and multilayered nature of our affiliations, Hegel thinks, further helps us to break away from the urge to reduce all reasons to individualistic ones. Members of the modern social world are not only citizens, but family members, members of various corporate entities associated with civil society, moral subjects and rights-holders. The distinctive value of modern social life lies, according to Hegel, in its ability to allow us to occupy each of these identities at the same time and bring them into a sort of harmony.[17]

---

without adding in this further wrinkle. I do, however, think that the model I develop below has the resources to address this problem. For further discussion of this point, see my *Reasonably Radical* (Ithaca: Cornell, 2001), especially chapter 6. It may help readers to see why there exist such resources to say that the Hegelian view developed below counts, as far as I can see, as what O'Neill calls a critical model of practical reasoning, rather than a norm-based one, and thus more like a view of practical reasoning she associates with Kant than one she associates with Hegel. See O'Neill, "Four Models of Practical Reason."

[17] See Hegel, *Elements of the Philosophy of Right*, §32. This harmony is one of the things that makes the modern social world worthy of what Hegel calls "reconciliation." For a discussion of the importance of reconciliation to Hegel's work, and the argument that the modern social world is worthy of reconciliation, see Michael Hardimon, *Hegel's Social Philosophy: The Project of Reconciliation* (New York: Cambridge, 1994). Note that the project of reconciliation involves showing each member of modern

For my purposes in this paper, the important lesson to learn is that the evaluation of social reasons is mischaracterized if it is presented, as the Hobbesian implicitly does, as a question of the relationship between social and individual reasons. We occupy a complicated web of social roles and the sets of social reasons that thus have a prima facie claim on us have a similarly complex geography. The effect of recognizing this complexity is that it changes the status of the individual standpoint. As the example of the European Parliament shows, sometimes evaluating the force of a given social reason involves evaluating it not from an individual standpoint, but from the standpoint of other plural subjects of which I am a member. We must therefore sometimes evaluate social reasons that appeal to one of our roles from the standpoint of a different one. Such perspectival evaluation provides standpoints that are external to any given plural subject without reducing the evaluation of all social reasons to a particular privileged standpoint.

I. C. The Space of Reasons à la Hobbes and à la Hegel. Having contrasted the two approaches along the structural dimension, I now argue that the Hegelian approach along this dimension offers a better approach to the evaluation of social reasons. The basic claim is that the Hegelian approach is theoretically better equipped to track the space of practical reasons, to keep us mindful of all and only the social reasons there are.

There are two requirements that an adequate evaluative standpoint must meet. It must be relatively stable and it must be relatively external. Many of the plural subjects that generate social reasons involve ongoing activities and play a role in structuring our lives over extended periods of time. If our participation in them is to have a positive impact on our lives, it must not be subject to short-term fluctuations in our allegiances. If, however, our evaluative standpoint were unstable, our evaluation of these practices would be subject to change even when the practices themselves did not change. Such change would undermine the relative stability of our allegiances, and this, in turn, would undermine the ability of these memberships to add value to our lives. I cannot realize the value of friendship in my life if I can only be a fickle friend. If the standpoint from which I evaluate my friendships is unstable, however, then this is the only sort of friend I can be.

---

society that they can comfortably occupy their various plural subjects, that these practical identities fit together, rather than showing, as the Hobbesian would aim to do, that they are each mutually advantageous.

The second requirement follows, in part, from the first. Since many of the plural subjects that provide us with social reasons serve to generate new reasons for us, our evaluative standpoint can only be stable if it is external to those practices. In addition, an evaluative standpoint that is not external will allow for a kind of circularity or bootstrapping. That certain practices are compelling or attractive for their participants once they have begun them does not guarantee that their participants have genuinely good reasons to continue them. Cults and addictions may generate internally self-supporting demands, but it is precisely in order to demonstrate the reasonableness of rejecting such claims that we need an evaluative standpoint in the first place.

The Hobbesian approach insists that an agent evaluate her attachments to plural subjects from her individual standpoint. While her individual standpoint is both stable and external, such a standpoint is not the only one that meets the requirements discussed above. In both cases, the Hobbesian relies on a particularly strong reading of the requirements, and in so doing, rules out the appeal of a number of social reasons whose force should not be so quickly dismissed.

The requirement of stability does not require absolute fixity, but rather something more akin to inertia. We can understand the need for our evaluative standpoint to have its own inertia in a strong or a weak form. On the strong reading, it must remain constant across the agent's becoming a member of the plural subject whose reasons are in question. Although it may be the case that I can come to have reasons I did not anticipate as a result of becoming a friend or a citizen, it must be the case that the reasons for continuing to be a friend or a citizen remain reasons that can be assessed independently of these roles. One consequence of this strong construal of the stability requirement is that my reasons for continuing to be a member of a plural subject and the reasons for my becoming a member in the first place must be drawn from the same set of reasons that were available to me at the time I joined the plural subject. It means, as I said above, that the Hobbesian cannot accept that my membership in a plural subject could generate new *types* of reasons for me. This is evident in the ease with which the Hobbesian approach adopts the language of contract in discussing the force of social reasons. Contracts change our obligations without changing us or giving us new types of reasons.[18]

---

[18] Hegel makes precisely this point about contracts; see *Elements of the Philosophy of Right*, §77. It forms one of the grounds of his criticism of social contract theory; see §75.

But the stability requirement need not rule out the transformative effects of being a member of a given plural subject. It is consistent with a weak reading of the stability requirement that my evaluative standpoint change over time as long as it does so relatively slowly and nonarbitrarily. Nothing requires that an evaluative standpoint remain constant across my becoming a member of a particular plural subject. Thus, for instance, George's reasons for becoming Alice's friend may be very different from his reasons for continuing to be her friend. The friendship may transform each of them over time, such that the point of view from which they each decided to become friends is no longer the appropriate one from which to evaluate the value of their friendship. Such a shift in their evaluative framework need not undermine their capacity for allegiance. In fact, it may be necessary for them to have that capacity at all.[19] In such cases, then, the Hobbesian approach may fail to appreciate these new types of reasons for maintaining a friendship, reasons that are generated as a result of the friendship and are not visible from the individual standpoint. If these new reasons also ground new social reasons within the friendship, then the Hobbesian approach will fail to evaluate properly the social reasons we have.

The Hegelian approach, on the other hand, satisfies the weaker form of the stability requirement. Within the Hegelian approach, there are a multiplicity of evaluative standpoints from which we assess our social reasons: the wide variety of relationships and memberships that go into making up our multifaceted identities. Their collective inertia is greater than the inertia of any one, and yet many of them also have their own inertia. Such inertia arises from the very structure of the plural subjects themselves, many of which have internal norms that determine how they change over time. These are the functional equivalents of voting rules and amendment procedures in constitutions that constrain how a state can adopt new directions and new constitutional provisions. As a result, the Hegelian approach can admit new types of reasons as grounding the normative authority of social reasons and thereby rule out fewer potential reasons. In doing so, it does not *require* that we admit the force of such considerations, but merely opens up the possibility that they count as reasons.

Turning now to the requirement that the evaluative standpoint be external, we can understand this requirement in both a strong and a weak fashion as well. The Hobbesian approach construes this requirement as demanding that the evaluative standpoint be not merely external to the given plural subject whose reasons are being assessed, but to all plural subjects of which the agent is a member. Once again, the individual standpoint satisfies this strong form of the requirement, but in doing so, rules out reasons that it should not. All that is necessary to avoid the problems of circularity and bootstrapping is that we can take up an evaluative standpoint that is external to the plural subject whose social reasons are being evaluated. This requirement can be satisfied by taking up a standpoint of a different plural subject. Thus, for instance, I can evaluate the reasons for friendship from any number of other plural subjects that I inhabit: from my perspective as citizen, moral agent, or family member. These will, alone or in combination, provide sufficient evaluative distance to avoid the problems that internal standards might run into. But they do not require that we take up, in addition to our various social standpoints, an individual standpoint outside all our substantive ties. Once again, the requirement that we do so places the bar for the evaluation of social reasons too high.

---

[19] Presumably, this sort of transformation explains Hegel's remark that "the precise nature of marriage is to begin from the point of view of contract...in order to supersede it"—*Elements of the Philosophy of Right*, §163.

The Hegelian approach, I have argued, provides us with an evaluative standard that, albeit complex and multiple, meets the requirements of relative stability and externality. It also rules out fewer possible social reasons at the outset. This less restrictive nature is an advantage. To see why, it is important to distinguish two means of rejecting purported reasons. First, we might try to reject the claims of others on prior theoretical grounds by arguing that their claims are not the kinds of things that could be reasons for us. Second, we might accept the possibility that these claims can count as reasons for us, and yet reject their purported authority in this instance on more substantive, concrete grounds. An approach to the evaluation of social reasons needs to provide restrictions on what counts as an adequate social reason, and so permissiveness is not, on its own, a virtue. Nevertheless, it is important that the restrictive quality of the approach take place at the right level. Both the Hobbesian and the Hegelian approaches provide grounds for the reasonable rejection of social reasons. The difference lies in where they do their work. The Hobbesian approach, I have been suggesting, frontloads its restrictions. The danger of doing so is that the theory blinds us to the possible force of whole categories of social reasons, which in turn impoverishes our sense of what reasoning together can be and do. Because the Hegelian approach is more open at the first stage, it is less likely to be blind. Because it provides adequate restrictions at the later stage, its greater capacity for vision need not dull its critical edge. It is thus likely to be a better guide to navigating properly the space of social reasons.

A sophisticated Hobbesian might reply at this point that nothing that I have attributed to the Hegelian view is beyond her reach. If we start from a sufficiently broad conception of what can count as in an individual's interests, then, the Hobbesian will claim, the individual standpoint will be able to take into consideration all the points raised by the Hegelian. Furthermore, to the extent that no sense can be made of membership in a plural subject being good for the individual, then we ought not countenance the demands such memberships make on us. In other words, a sophisticated Hobbesian can help herself to everything attractive in the Hegelian structure, and nothing she cannot help herself to is worth having.

There is certainly room for such a Hobbesian position. We can certainly understand that as individuals we have interests in forging and maintaining all sorts of substantial social ties, and in partaking in what Charles Taylor calls "irreducibly social goods." Since, however, the value of such things is ultimately for individuals, it does look as if they should be evaluated from that standpoint and not any other.[20]

While the sophisticated Hobbesian might thus take advantage of many of the Hegelian insights and shifts I have been laying out, there is one aspect of the Hegelian picture that I think the Hobbesian is not in a position to take on board. For the Hobbesian view to have the determinate content that is its great attraction, it must begin with a settled account of the sources of evaluation available from the individual standpoint, so that that standpoint can provide a fixed point from which the evaluation of our membership in plural subjects can take place.[21] The consequence of this requirement is that the reasons for being a member of a plural subject will have to be restricted on the Hobbesian view, since they will always have to point back to already set factors available from the individual standpoint. The Hegelian view, on the other hand, can adequately capture how our evaluation of social reasons can change as a result of changes in our understanding of the reasons for being a member of various plural subjects in the first place, and the fact that such changes can take place through our collective deliberation. In particular, it can describe properly cases where our continued participation in a plural subject, including deliberation within it, generate altogether new grounds for maintaining our membership in it.[22]

---

[20] Taylor would, I think, deny this point. He claims that the goodness of an irreducibly social good requires not only our common understanding of the good, but our common valuation of it as good. See "Irreducibly Social Goods," p. 139. The Hegelian position described above differs from Taylor on this point. According to the view I am articulating here, whether the evaluation of social reasons can devolve back to individual members of the plural subject depends on the particular content and constitution of the plural subject, not merely on its being plural. This view is thus vulnerable to the sophisticated Hobbesian objection in a way that Taylor's view might not be. On the other hand, it is also thereby a position a Hobbesian might be more likely to find attractive.

[21] We need to understand this claim at the right level of generality. Of course, there is nothing that prevents the Hobbesian from saying that an individual's tastes or preferences change and that these changes change the reasons she has. What the Hobbesian must insist on is that whatever aspects of an individual ground reasons, this set must remain fixed. If, for instance, preferences are the only ultimate source of reasons, then it is a fixed point of the theory that preferences remain the only ultimate ground of reasons.

According to the Hegelian approach, we can, in the course of our deliberations, come to grasp new types of reasons for membership and this can change the nature of our belonging and this can further affect how we evaluate reasons elsewhere. Thus, even if the status of given social reasons depends on our reasons for being a member of the plural subject in question, it may turn out that our reasons for being a member now are not the same as our reasons for becoming a member in the first place. The European Union provides a particularly vivid example of this kind of transformation. In the course of developing the institutions of the Union, the member nations and their citizens have also developed an evolving understanding of what the Union is for. Originally conceived as a way of establishing a lasting peace in Europe in the wake of two world wars, the union evolved into primarily an economic bloc, concerned not so much with maintaining internal peace as with competing in the global market. As it now moves beyond basic economic integration, one of the motivating factors for European membership that is beginning to surface is the protection of various sub-European identities (whether national or sub- or trans-national). As these self-understandings change, the kinds of reasons that stem from membership in the EU change also, and with them, the relevant place of these reasons in the reasoning of particular European citizens.

Note that this difference between Hobbesian and Hegelian approaches parallels a difference I noted above between the sub-Hobbesian and the Hobbesian view. The sub-Hobbesian can admit new reasons, but not new kinds of reasons into an individual agent's standpoint. As George becomes Alice's friend, her well-being affects his own. If George's well-being gives him reasons to act on the sub-Hobbesian view, then he now also has reasons to promote Alice's well-being. But these are reasons that find support in his well-being, not hers, and so are not a new type of reason for him. The Hobbesian, however, accepts that reasons of friendship can be genuinely new types of reasons for George. At the same time, she insists that such new considerations can only be reasons if the plural subject that supports them can appeal to George's individual standpoint. Thus, there can be no new types of reasons for membership in a plural subject. Once again, this does not require the Hobbesian to deny that there can be new reasons for membership, just new types of reasons. Thus, George might initially become Alice's friend because of the fun they have together, and remain her friend because of the access she provides him to a social world he hopes will further his career. But both reasons are of a kind: each appeals to George's original individual standpoint. My claim is that the Hegelian approach allows us to recognize the possibility of new kinds of reasons for remaining within a plural subject, reasons that appeal to different evaluative standpoints, for instance. A Basque manufacturer, for instance, might have originally found grounds for supporting European membership in his individual interest in not fighting in a war, but come to find new reasons stemming from the advantages available to his firm, and finally the possibility of protection for Basque culture and language. Since all of these standpoints are themselves capable of being transformed by their European context, it is possible for new kinds of reasons to develop whose development cannot be explained independently of that context.

---

[22] This also marks a difference with Taylor's view, as I understand it. For Taylor, practical reason involves a kind of interpretational discovery rather than a kind of mutual (re)-construction. See, for instance, "Explanation and Practical Reason," p. 36. The claim that there is a fact of the matter about what practical reasons there are that is prior to the act of reasoning together, then, is a point on which Taylor agrees with the Hobbesian and not the Hegelian.

II. SUBSTANTIVE DIFFERENCES

Though the reply above to the Hobbesian objection points to a virtue of the Hegelian position, it is unlikely to calm the concerns of the Hobbesian. Those concerns require us to look into the second dimension of difference between the two approaches. Here, the difference between the two approaches is more pronounced. The Hobbesian relies on a theory of practical reason that takes the ingredients of practical reasons to be individual mental entities, such as desires, preferences, or interests. The Hegelian, on the other hand, takes the ingredients of a practical reason to be principles of the will.[23] In some sense, this difference mirrors the structural differences, insofar as the Hobbesian regards reasons as essentially private, whereas the Hegelian regards them as essentially public. Moreover, it is, to a large degree, the substantive commitments of the Hobbesian that push her to hold on to her structural commitments and prevent her from seeing the important structural differences between the two approaches.

---

[23] Much philosophical discussion that distinguishes Hobbesian from Kantian or Rousseauvian approaches to practical reason has focused on this substantive difference. See, for instance, O'Neill, "Four Models of Practical Reasoning"; Barbara Herman, *The Practice of Moral Judgment* (Cambridge: Harvard, 1993).

On the Hobbesian approach, practical reasons are built up out of psychological entities that are not themselves necessarily rational. That is, the various sorts of psychological entities that go into practical reasons are entities that one could conceivably find in creatures without practical reason. Thus, for instance, on a belief-desire model of practical reason, a practical reason is made up of a belief and a desire, but it is also the case that creatures who do not reason have beliefs and desires. This holds even for views that build reasons out of more sophisticated psychological entities like intentions, preferences, or interests.

Since the Hobbesian assumes that practical reasons must be built out of such materials, he assumes that the Hegelian approach to practical reason must also rely on these materials. The problem, however, is that if these materials are the properties of individual minds, it is hard to see how there could be reasons that are not accessible from an individual standpoint. Thus, the Hobbesian's substantive position informs his understanding of the structural differences between the two approaches, and makes it seem as if the Hegelian position is either nothing new or mysterious. The Hegelian position is nothing new if the reasons available from a social evaluative standpoint must nevertheless be built out of the same psychological entities as those that inform the individual standpoint. The Hegelian position will be mysterious if the reasons available from the social evaluative standpoint are built out of an altogether new kind of entity, such as a platonic reason or a collective interest.

Take, for instance, my earlier claim that on the Hegelian approach, my evaluation of my being a father cannot be appropriately undertaken from an individual standpoint, along with my acknowledgment that I can nevertheless ask, as a father, whether being a father is good for me. If we hold onto the substantive Hobbesian position, it is hard to make sense of this claim. One possible reading of the claim is that being a father is good for me because I have reasons to continue being a father, and these are the same sort of reasons that I have to do or be anything else: they are built up out of my individual psychological furniture. Being a father satisfies certain desires, or fulfills certain interests. Of course, these may be desires and interests I only came to have upon becoming a father. But nothing here is out of reach of the Hobbesian approach. Such factors would not generate new types of reasons.

Thus, if the Hegelian is really saying something new, she must take my reasons to be a father to be of a new sort (that, after all, is what she says). They must, then, be reasons that are built out of a social evaluative standpoint. This suggests that they must come from something like collective psychological entities, like collective interests, desires, or beliefs. Alternatively, they may come from some extra-psychological source, and have application only to fathers or family members and not to unattached individuals. At this point, however, the Hobbesian is likely to ask just what are these new mysterious entities that are not just reducible to their individual counterparts.

David Gauthier, in considering alternatives to the Hobbesian approach, makes both of these moves. In *Morals by Agreement* (*op. cit.*), Gauthier contrasts the Hobbesian approach to reasoning together with what he calls the universalistic conception of reasons. The universalistic approach claims that the authority of reasons is an objective matter, unconnected to the interests of the particular agent to whom the reason is addressed. If my being a citizen of a given country gives me reasons to act, the authority of these reasons lies in the objective goodness of the state. Similarly, if my mere humanity provides me with reasons, this has to do with the objective properties of the moral law, quite apart from any interests I do or ought to have. Gauthier argues in favor of working from the Hobbesian approach because he takes its assumptions about the normativity of reasons to be weaker than those of the universalistic approach.

Elsewhere in his writings, Gauthier considers a third approach to the authority of reasons that he traces to Rousseau and Hegel.[24] This approach ties the authority of reasons to the interests of collectives of which agents are members, and so moves beyond the individualism of the Hobbesian approach while maintaining some connection between the authority of reasons and individual, prerational, psychological elements. Gauthier's regret at the rejection of this approach is more palpable. He regards this theory of practical rationality as more attractive, but ultimately false, either to deep-seated facts about human nature, to our own current self-conceptions, or to the contractarian ideology that governs how we approach questions of reasoning together here and now.[25]

---

[24] The core of this idea in Rousseau is described in Gauthier, "The Politics of Redemption," *Revue de l'Université d'Ottawa*, XLIX (1979): 329–56, reprinted in *Moral Dealing: Contract, Ethics, and Reason* (Ithaca: Cornell, 1990), pp. 77–109, especially pp. 92–94. The general approach, as well as its connection to Hegel, forms the backdrop of Gauthier, "Social Contract as Ideology," *Philosophy and Public Affairs*, VI, 1 (Winter 1977): 130–64, reprinted in *Moral Dealing*, pp. 325–54.

In both cases, Gauthier seems moved by a worry that the alternative approaches fail to capture adequately the close connection between an individual agent and her reasons. Universalistic reasons are free-floating, and so it can seem mysterious how they are ever supposed to move us to act. Reasons of collective interest seem, on the surface, to be in a better position to answer this worry, but in the end run into similar problems. If collective interests are not reducible to the interests of individual members of the collective, then they must be properly the interests of a separate interest-bearing entity: the collective. But then the connection between an individual and the interests of the collectives of which she is a member will also be mysterious.

The Hegelian approach avoids these problems, however, because of its substantive position about the building blocks of practical reasons. It relies on a theory of practical reason that grounds practical reasons in principles of the will.[26] On this view, which has roots in Rousseau and Kant, practical reasons are the result of features unique to rational agents. To have practical reason, according to this view, is to have the ability to determine one's action in line with rational principles. Practical reasons, then, are built up from rational principles, not from individual psychological entities. These principles, however, have the authority they do for us because of how our wills are constituted. In other words, they have the authority they do for us because of who we are. Thus, insofar as our membership in some groups provides us with certain principles, our membership gives rise to reasons directly. What it is to be a member of a plural subject is to have one's practical identity in part constituted as a member, and that means to take the norms of that plural subject as authoritative, as giving rise to reasons.

Thinking of practical reasons in terms of principles rather than individual psychological entities makes reasons essentially public without rendering their connection to individuals mysterious. If a rational principle applies to both of us, then we share that principle and the reasons it authorizes. Contrast this with the sense in which if we each have a certain belief and desire, say, we each have a similar reason. In the latter case, there is a sense in which we do not really share the reason, although we each have it. Because social reasons are shared, they are of a sort not available to the individual evaluative standpoint in the same way. It is this public quality of practical reasons on the Hegelian view that makes its insistence on the irreducibility of the social evaluative standpoint importantly different from the Hobbesian approach.

---

[25] Inconsistency with a deep fact of human nature seems to be what sinks Rousseau in "The Politics of Redemption." Inconsistency with our "everyday awareness of ourselves" prevents this theory from providing a reconciliation between rationality and Rawls's theory of justice in Gauthier, "Justice and Natural Endowment," *Social Theory and Practice*, III (1974): 3–26, reprinted in *Moral Dealing*, pp. 150–70. Failure to find support in our ideology is discussed in "Social Contract as Ideology."

[26] O'Neill draws a similar contrast between two broad types of theory of practical reason, one that she describes in terms of "preference-based" versus "action-based" theories. See her "Four Models of Practical Reason."

At the same time, to say that practical reasons must be grounded in principles of the will is not to ground them in metaphysical facts external to the individual for whom they are reasons. Rather, it is to ground practical reasons in a different part of our psychology. Principles that ground practical reasons ground those reasons for me only when I am suitably attached to those principles. This aspect of the view is captured in the claim that the norms of a plural subject that constitute principles for my will serve thereby to constitute part of my practical identity. What counts as suitable attachment is a complicated question that goes well beyond the scope of this paper, and so I leave that question open here. Nevertheless, to claim that practical reasons are built out of our practical identities rather than our desires or interests does not separate reasons from the individual in the way the universalistic and collective interest views do.

In addition, although principles can be shared, and this is one of the central means through which we come to form and constitute plural subjects, shared principles are to be distinguished from collective interests. In fact, one of the advantages of shifting from an interest-based to a principle-based theory of practical reason is precisely that it provides the means for conceiving of a group of people sharing reasons that nevertheless appeal to each individually, without having to appeal to each person's prior individual interests. If reasons find their support in interests, then we need to posit collective interests in order to ground shared reasons, and this leads to the difficulties that Gauthier sees with the collective interest approach. By shifting to a principle-based theory of practical reason, we have the resources to understand how reasons can be shared without having to posit new and somewhat mysterious collective entities with interests of their own. As a father, my practical identity supports certain principles and these give me social reasons to act. If I now ask what reasons I have to take these reasons seriously, the answer can only come from rational principles that have authority for me. These will include, but not be limited to, the principles that have authority for me insofar as I am a father. From the point of view of a principle-based theory of practical reason, there is nothing mysterious here, and yet, there is more going on than an evaluation of reasons from an individual standpoint.

The Hobbesian might reply, however, that this move merely sweeps the mystery under the rug of practical identity. That is, we are now left with the question of how to explain the connection between my practical identity and my action. Given that practical identity has not been precisely defined here, this connection may seem no less mysterious than the one that sunk the collective interest approach in the eyes of the Hobbesian. It is undeniable that practical identity, understood as a psychological concept, is more complex and hard to define than such psychological concepts as interest, desire, or end. Nor have I done any serious work in this paper in trying to provide a precise psychological account of practical identity. Such work, however, is not needed, and it is the mistake of this objection to think that it is. The problem with this objection is that it focuses on the wrong set of concepts. The terms that lie at the foundation of the Hobbesian approach to reasoning together, like "desire," "interest," or "end," are used in two different ways that are not often clearly distinguished.[27] Sometimes they are used psychologically, to refer to a mental state of some sort. In other cases, they are used as terms in a theory of practical reasoning. Within a given theory, they denote the kinds of objects that make up part of a practical reason. In this latter use, their definition depends entirely on the functional role they play within the theory as a whole. Thus, something counts as a desire in this latter sense, say, not because of the way it feels to the one who has it, but because it provides her with the basis of a reason to act, according to the operative theory of practical reason. That it is the connection to the space of reasons and not to our psychological states that determines what counts as a desire in this latter sense is often marked off by distinguishing desires from cravings or mere urges. Cravings and mere urges *feel* like desires, but do not connect up to the space of practical reasons; they are, we might say, normatively inert, even if psychologically powerful.

The importance of this distinction between normative and psychological uses of the same term is that the precision of the normative term is in no way determined by the precision of the corresponding psychological term. Thus, turning back to the difference between the Hobbesian and Hegelian approaches, no support for the Hobbesian view can be found in our greater comfort with and precision in our understanding of the psychological use of terms like "interests" and "desires" than with the psychological use of terms like "identity." Rather, what is relevant to the comparison is whether or not the normative use of practical identity is sufficiently well defined. That question, in turn, depends on the coherence of each term's position within their respective theories of practical reason, and here, I would submit, there is no significant difference to be found. Just as "interest" names that which can ground a practical reason at the foundational level for the Hobbesian, "practical identity" names that which can ground a practical reason at a fundamental level for the Hegelian. It is certainly the case that the two types of theory are substantively different, and that the content of Hegelian theories will be more complex. But, given the argument above, these facts do not imply that one is less well defined than the other.

---

[27] One exemplary exception to this, whose argument I follow below, is Elijah Millgram, *Practical Induction* (Cambridge: Harvard, 1997), especially chapter 2.

III. EVALUATING SOCIAL REASONS, HEGELIAN STYLE

Having discussed at length the two axes on which the two approaches differ, we are now in a position to construct a Hegelian approach to the evaluation of social reasons. Such evaluation can take one of two intertwined forms. The first is internal to the structure of the relevant plural subject. It evaluates the reason offered in terms of the point of the plural subject in question. Although one possible point for the continued existence of a plural subject is that it provides for the mutual advantage of the parties, this is not the only grounds for forming and maintaining plural subjects. We might think about the problems of human coexistence that certain relationships, institutions, and groups help to meet and how they do this. Such reasoning will involve a great deal of interpretation, as the point of our various memberships or how these translate into reasons may not always be obvious. To give an example, what reasons we have in virtue of our being citizens may very well depend on how we conceive of our political relationship as citizens, and its point. Is it primarily a means of avoiding violent death at other's hands, as Hobbes seems to have believed? Then we will have reason to endorse any form of regulation that secures this end. If, however, the problem that political organization solves is that we continuously run into prisoner's dilemmas and other problems of collective action given our rationality, then we will have reason to support regulation that brings about something like Gauthier's morals by agreement. If, finally, the problem we face is how to respect our differences and not impose our reasons on others who reasonably disagree with us, then we will have reasons for adopting some form of what Rawls calls political liberalism.

In addition to this internal, interpretive investigation, we can also evaluate social reasons by thinking about how the claims of the plural subject that ground a given reason fit together with our membership in other plural subjects, and the claims those make on us. Thus, for instance, if I am to evaluate the kinds of obligations and opportunities that being a member of a family places on me, I will do this not only by asking how it impacts my ability to pursue my pre-existing individual interests, but also how it impacts my ability to live as a member of various other groups. I might reasonably object to the claims of family if they unduly interfere with my ability to be an active citizen or pursue a professional career, just as members of my family might reasonably object to my pursuit of my career or my political activities if they unduly interfere with my capacity to be a good father and husband. Thus, one of the grounds we may have for rejecting a given social reason and the conception of the plural subject on which it rests is that it leaves insufficient room for our other allegiances and memberships. In evaluating reasons in this way, no particular standpoint plays the fundamental role played by the individual standpoint on the Hobbesian approach. Rather, we must balance and adjust, examining each membership from the standpoint of others. Does my being a citizen of a state like this make it overly burdensome to be a member of a family or a certain religion as well?[28] If so, then perhaps I need to reinterpret one of my allegiances so that they fit together more comfortably.[29] If I cannot do that, then I may have grounds for rejecting the social reason that created the tension, or in extreme cases, the aspect of my identity that gave rise to it. That my rejection of such reasons would be reasonable then provides grounds for my fellow members to reconsider the feature of our plural subject that supported the claim in question. However we finally resolve this tension, we will be led to greater clarity by reflecting, as on the first path, on the reasons each of us has for being a member of each of the plural subjects in tension. Notice that as with the Hobbesian approach, my acceptance or rejection of a social reason can be traced to my reasons for being a member of the plural subject that supports it. Unlike on the Hobbesian approach, however, my reasons for membership can be much broader.

---

[28] I discuss the place of such reasons in political deliberation in my *Reasonably Radical*, chapter 7.

[29] I should note at this point a key feature of the Hegelian approach that departs from Hegel. For Hegel, the state occupies a pre-eminent place among the institutions of the modern world, and so there is an important sense in which conflicts between my allegiance to the state and other plural subjects would not, according to Hegel, be resolved by balancing, but by submission to the dictates of the state. Nevertheless, one of Hegel's justifications for giving the state such a pre-eminent place was that it did make room for its citizens to take up the various other identities he saw as essential to the realization of our freedom, so while active balancing may not fit into Hegel's account, a concern with the harmonious fit of varying identities and allegiances certainly does. I am grateful to Isaac Balbus and Stephen Engelmann for pushing me to be clearer on this point.

Clearly these two forms of evaluation intertwine. Tension between the demands of two groups or institutions of which I am a member will push me to reconsider how I conceive of them, and the demands they place on me. Reinterpreting a particular membership will impact on how that membership can coexist with others that I, or others, might value. Thus, on the Hegelian approach, there may be no neat theoretical answers to the question of what force a given claim really has for us. Faced with a given claim, we may need to think about the point of the plural subject that supports it, and the relationships amongst the other plural subjects to which we belong, and each of these considerations will be related to the others. Furthermore, in most cases, it is not up to individual members to reformulate unilaterally the obligations membership in a plural subject place on them, or even the nature of their affiliation. Thus, pressure to reinterpret a given plural subject of which I am a member can only lead to my making demands on others to join in such a project. Since there is no fixed set of grounds for evaluating reasons, it will be impossible to develop a straightforward theory of reason evaluation similar to a theory of rational choice that derives the authority of certain reasons from a set of allowable functions on a fixed set of starting points.

How, then, are conflicts between different social reasons to be resolved? The short answer is deliberatively. We need to weigh a variety of considerations with other people, and be attentive to the various kinds of demands our various memberships place on us and them, and the variety of reasons we have for affirming those memberships in the first place. Thus, for instance, the degree to which I will let my political membership as a citizen of a given state infringe on my ability to be a practicing member of a particular religion may well depend on how I conceive of each affiliation and the grounds I have for continuing them. The evaluation of social reasons thus turns out to require a search for a kind of reflective equilibrium between not only the demands of the various plural subjects to which we belong, but also the conceptions we have of those plural subjects. That is, there is no longer a neat divide between what might be called constitutional deliberation, wherein we work out the nature of our shared identity, and executive deliberation, where we work out what the content of that identity commits us to doing. Figuring out what to do will not be an activity wholly separate from figuring out who we are, and neither will be a question of deriving answers from a set of static premises. In addition, because who we are is in large part determined by our membership in various plural subjects, and it is in general not up to any one of us to change their content unilaterally, our reasoning together will have to aim not only at the internal equilibrium of each, but at the joint equilibrium of all.[30]

IV. THE WORK AHEAD: WHERE DOES IT LIE?

I have contrasted two approaches to evaluating social reasons with the aim of making the Hegelian approach look both attractive and distinctive. I have raised and responded to a number of Hobbesian objections along the way. There are, no doubt, many others that I have not considered. Most Hobbesian objections to a Hegelian approach share a basic philosophical orientation, however, and in this final section, I bring out this fundamental difference and offer some general reflections about it.

The Hobbesian approach rests firmly on an individualist attitude that a proper theoretical parsimony requires that a theory of human reason and behavior start from basic building blocks that can be found *within* individuals, and work outward, adding new blocks only when the combination of the ones we have already got cannot adequately explain something in need of explanation. Thus, we should start with desires and beliefs, and perhaps add in intentions or plans, and so forth. On this view, the problem with the Hegelian approach is that it too quickly jumps to introduce extra, potentially unnecessary new blocks, blocks that are themselves often somewhat mysterious. This criticism misses the structure of the Hegelian approach, however. The Hegelian theory need not be any less theoretically parsimonious than the Hobbesian approach. It merely starts in a different place, with a different set of blocks. Rather than starting within individuals, and trying to work out to the social, the Hegelian approach starts with intrinsically social entities (social reasons and plural subjects) and works inward, giving it the task of adequately explaining the very individualist blocks from which the Hobbesian starts. As a result, the theoretical tasks it faces will be different

At first glance, this claim will strike the Hobbesian as absurd. How can the explanation of our very individuality be a philosophical task in the way that the explanation of our ability to interact with others can? Surely any approach that does not take very seriously the fact that human beings are, fundamentally, individuals should for that reason alone be rejected.

---

[30] It is this aspect of the Hegelian view that makes it a form of what O'Neill calls a critical conception of practical reason.

The Hegelian approach, however, no less than the Hobbesian, accepts this fact. Furthermore, it accepts that this fact has deep and far-reaching implications for any theory of practical reasoning. Where the Hegelian approach differs, I suggest, is in the role that it assigns to the fact of our being individuals in the theory of practical reasoning and thus the theoretical work it faces.

The Hobbesian approach includes the fact of our being individuals in its picture of practical reasoning by stipulating that the mental furniture of our individual psyches must provide the building blocks of reasons, and our individual standpoints must act as a foundational evaluative perspective. If there are to be social reasons at all, the Hobbesian claims, they must derive their ultimate authority from their connection with these aspects of individuals. In this way, the Hobbesian places our being individuals at the foundation of the theory of practical reasoning. It also thereby sets itself the problem of providing an adequate account of the role of social reasons in reasoning together.

The Hegelian approach starts from a different point, with the normativity of social reasons and the foundational place of our memberships and relationships in our reason-guided agency. From the Hobbesian perspective, it thus looks as if the Hegelian fails to take sufficiently seriously our being individuals. But the Hegelian takes this fact seriously, as something the theory must explain rather than a premise from which it must start. This, indeed, sets the Hegelian a theoretical task: explaining the undeniable importance of our being individuals. How might this work proceed? On the Hegelian approach I have outlined here, the fact of our being individuals enters the picture of practical reasoning in two ways. First, one of the many ways in which we conceive of ourselves will be as individuals. Such a form of self-conception, however, is not, according to the Hegelian, a kind of natural outgrowth of our being biologically distinct, but rather a result of particular social organizations. What it means to be an individual will depend on many other aspects of our self-conceptions and those of our society: seeing myself as having a soul, or being made in the image of God, or as a rational agent of economic theory or as a rational agent of Kantian theory are all ways of conceiving of myself as an individual, but they are all very different, and may yield a rather different array of reasons.[31] The Hegelian approach thus allows for individual standpoints, so conceived, to play a role as one of the many standpoints from which we can evaluate social reasons.

---

[31] For an illuminating discussion of how this point plays out in Hegel's work, see Terry Pinkard, *Hegel's Phenomenology: The Sociality of Reason* (New York: Cambridge, 1994).

More interestingly, however, the Hegelian approach offers another place for the fact of our being individuals to play a role in the theory, one which gives this fact a kind of foundational importance that it lacks in the role described above. According to the Hegelian approach, our being individuals places constraints on the collection of reasons we can face. Thus, for instance, part of what allows me to reject certain social reasons offered to me in my role as member of a certain group is that if my membership in that group authorizes reasons like that, then it will be impossible for me also to inhabit another role that I also value. Such rejection rests on what I called balancing considerations above. But notice, now, that the force of balancing considerations turns on my limitations as an individual. That is, they turn on such basic facts as my need for a kind of affiliational coherence, because I am incapable of affirming two contradictory statements, or taking them as the basis of my actions or self-conception, or on my inability to be in two places at once, or to do activities that require four hands. Thus, on the Hegelian approach, the importance of the fact that I am an individual is, as it were, woven into the very structure of practical reasoning. The difference between the Hobbesian and the Hegelian approach, then, lies not in the importance each places on our being individuals, but the way they do so, and moreover, the degree to which each has theorized about the question. That is, while there has been much attention paid to the question of individual interests in the philosophical, economic, and psychological literature, much less work has been done to understand the details of being an individual from the point of view of the constraints it places on our practical reasoning.[32] That more work needs to be done on this question, however, is no reason to reject the Hegelian approach in favor of the Hobbesian one. It is merely a reason, if the foregoing arguments have merit, to roll up our sleeves and get to work.

ANTHONY SIMON LADEN

University of Illinois at Chicago

---

[32] It seems to me that there might be quite interesting constraints beyond the obvious "can't be in two places at once" variety, and that thinking about the relationship between practical reasoning and the individual self from this perspective might shed light on both practical reasoning and the self. For a taste of the kind of reflection and constraints I have in mind, see Marya Schechtman, *The Constitution of Selves* (Ithaca: Cornell, 1996), especially part II, where she develops what she calls a narrative view of the self.

IS THERE A SINGLE OBJECTIVE, EVOLUTIONARY TREE OF LIFE?*

It is often said that there is just one "objective" tree of life: a single accurate branching hierarchy of species reflecting order of descent. For any two species, there is a single correct answer as to whether one is a "daughter" of the other, whether the two are "sister species" by virtue of their descent from a common parental species, whether they belong to a family line that excludes any given third species, and so on.

The idea is intrinsically interesting. It has consequences for what we should think about the evolutionary origins of *Homo sapiens* as well as other species. The idea also has important connections to the scientific discipline of systematics, which classifies organisms into related groups. The connection to systematics is what has made objectivity a topic of discussion for biologists and philosophers of biology. Cladism, now the dominant school of systematics, places organisms into groups depending just on their place in the tree of life. That is supposed to render classification objective.

In section I, I recite claims to objectivity. In sections II–V, I argue that the apparent objectivity is not what it seems to be. In section VI, I briefly revisit systematics.

I. THE APPEAL TO OBJECTIVITY

Cladistic classification, founded by Willi Hennig, takes into account just relationships of ancestry and descent.[1] The alternative is to take into account evolved similarity. Similarity allows for a multiplicity of permissible classifications. Paul E. Griffiths and Kim Sterelny follow many others in stressing this: "as cladists never tire of pointing out—similarity depends on the traits you measure."[2] In some respects, chimpanzees seem more similar to gorillas than to humans. Both are hairy and share a relatively primitive lifestyle. In other respects more obvious to specialists than laypersons, chimpanzees are more similar to humans. Similarity might yield no unique hierarchical representation.

It is supposed to be an "axiom" of cladistic systematics that an unambiguous *historical* hierarchy *is* available for reference: "nature's

---

* For discussion, I am grateful to many, especially Bruce Aune, Brian Bodenbender, and James Walker.

[1] Hennig, *Phylogenetic Systematics* (Urbana: Illinois UP, 1966).

[2] *Sex and Death: An Introduction to Philosophy of Biology* (Chicago: University Press, 1999), p. 196.

0022-362X/05/0207/357–74

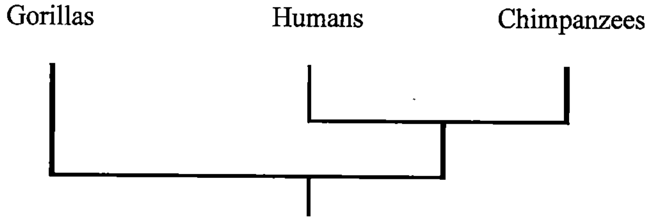


Figure 1: Humans and chimpanzees comprise a *monophyletic* group; gorillas and chimpanzees without humans do not. Monophyletic groups include all descendants of the closest common ancestral species.

hierarchy is discoverable and effectively represented by a branching diagram."[3] With respect to humans, chimpanzees, and gorillas,

> History has an objective structure. Suppose we share a more recent common ancestor with the chimp species than it shares with the gorilla species—that the ancestors of the gorillas diverged from a branch ancestral to both chimps and humans. If so, then historical phylogeny should put humans and chimps together in a more closely related group than any including both chimps and gorillas. If not, then it's just a mistake to think of humans and chimps together alone as a single group.[4]

As it happens, scientists are agreed that humans and chimps do share a common ancestor not shared by gorillas (see figure 1). Because that is all that we need to know to classify the three with respect to one another for cladism, we hear that "Cladistic classification has the advantage of objectivity. The phylogenetic hierarchy exists independently of the methods we use to discover it, and is unique and unambiguous in form."[5] Of course, in practice we may have a hard time *discerning* relationships. Still, there is one correct account of the relationships: "One True Tree of Life."[6] The state of our information is irrelevant because regardless of that, "It really is true that two species either do, or do not, share a more recent ancestor with each other than with any other species."[7]

---

[3] I quote from Robert W. Scotland, "Cladistic Theory," in Peter L. Forey, Christopher J. Humphries, Ian J. Kitching, Robert W. Scotland, Darrell J. Siebert, and David M. Williams, eds., *Cladistics: A Practical Course in Systematics* (New York: Oxford, 1992), pp. 3–13, at p. 3.

[4] Griffiths and Sterelny, *Sex and Death*, p. 197.

[5] Mark Ridley, *Evolution* (Malden, MA: Blackwell, 2004, 3rd ed.), p. 480.

[6] Griffiths and Sterelny, *Sex and Death*, p. 194.

[7] Ridley, "Can Classification Do Without Evolution?," in Ridley, ed., *Evolution* (New York: Oxford, 1997), pp. 196–206; this citation is from p. 202. That cladism "allows only one classification: the classification defined by the hierarchy of phylogenetic branching," as Ridley is always reminding his many readers, is widely accepted: *Evolution and Classification: The Reformation of Cladism* (New York: Longman, 1986), p. 46. See Griffiths, "Cladistic Classification and Functional Explanation," *Philosophy*

II. SPECIES AND SPECIES CONCEPTS

The alleged uniqueness and absence of ambiguity concerning the phylogenetic hierarchy is exaggerated. I will show in the following sections that different accounts of the familial relationships between species can be acceptable. The notorious species problem is enough to give legitimacy to distinct accounts. In the current section, I rehearse that problem, arguing that no one "species concept," or account of what a species is, can be discovered to be the correct one at the expense of others. In section III, I go on to argue that if the arguments from section II are sound, then a group of species may allow for many valid accounts of the order of evolutionary branching and the resulting familial relationships. In section IV, I show that appeal to the species problem is dispensable. There are many valid, competing accounts of evolutionary branching even if, contrary to what I argue in section II, there is a single valid species concept. In section V, I discuss the nature and extent of valid diversity in accounts of evolutionary relationships.

As I have indicated, the present section is devoted to the species problem. Biologists make heavy use of the concept or concepts of *specieshood*: they discuss the habits and movements of this or that species, the predators, prey, and parasites of a species, the number of species in an ecosystem, and so on. But there is little agreement about what species are. There are dozens of live species concepts.

Two popular concepts are Ernst Mayr's *biological species concept* and Joel Cracraft's *phylogenetic species concept*. According to the biological species concept, species are "groups of interbreeding natural popula-

---

of Science, LXI (1994): 206–27, at p. 207; David Hull, "The Role of Theories in Biological Systematics," *Studies in History and Philosophy of Science, Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, XXXII (2001): 221–38, at p. 224; "The Limits of Cladism," *Systematic Zoology*, XXVIII (1979): 416–40, at pp. 437–38; Samir Okasha, *Philosophy of Science: A Very Short Introduction* (New York: Oxford, 2002), pp. 111–12; Ridley, *Evolution*, pp. 475, 479, 481, 490; "The Cladistic Solution to the Species Problem," *Biology and Philosophy*, IV (1989): 1–16; *The Problems of Evolution* (New York: Oxford, 1985); Donn Eric Rosen, "Cladism or Gradism?: A Reply to Ernst Mayr," *Systematic Zoology*, XXIII (1974): 446–51, at pp. 447–48; David N. Stamos, *The Species Problem: Biological Species, Ontology, and the Metaphysics of Biology* (Lanham, MD: Lexington, 2003), p. 265; E.O. Wiley, *Phylogenetics: The Theory and Practice of Phylogenetic Systematics* (New York: John Wiley and Sons, 1981), p. 268. Foreshadows of the problems that I raise may be found, in some measure or other, in L.A.S. Johnson, "Rainbow's End: The Quest for an Optimal Taxonomy," *Systematic Zoology*, XIX (1970): 203–39, at p. 231; Robert O'Hara, "Systematic Generalization, Historical Fate, and the Species Problem," *Systematic Biology*, XLII (1993): 231–46, at pp. 232–33; and Sterelny, "The Nature of Species," *Philosophical Books*, XXXV (1994): pp. 9–20 at p. 17, who appears since to have had a change of heart (cf. Griffiths and Sterelny, *Sex and Death*, pp. 194–97): in section V, I indicate that there is something importantly right in each position.

tions that are reproductively isolated from other such groups."[8] According to the phylogenetic species concept, "A species is the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent."[9] These two concepts both divide the world into natural groups, but often the groups are not the same: the groups recognized by the biological species concept tend to be more inclusive than the narrower groups recognized by the phylogenetic species concept. The biological species concept may group different "races" or "subspecies" because they are interfertile, while the phylogenetic species concept distinguishes them into different "species." Hence, the biological species concept divides the birds as a whole into an established 10,000 species. The phylogenetic species concept would double that number to 20,000 species.[10]

If either the biological species concept or the phylogenetic species concept could be discovered to be the correct species concept, then this could spare witnesses of one unambiguous family tree of species from a source of embarrassment. But it seems hard to believe that either the biological species concept or the phylogenetic species concept could be discovered to be the correct species concept. The decision about whether to adopt the biological species concept or the phylogenetic species concept admits more than one acceptable resolution.

The choice between competing proposals is not completely arbitrary: were it not the case that the biological species concept and the phylogenetic species concept both divide the organic world into groups that are natural and scientifically interesting, there would be no competition between them. But each concept seems to divide the organic world into groups that are natural and scientifically interesting.

---

[8] Mayr, "The Biological Species Concept," in Quentin D. Wheeler and Rudolf Meier, eds., *Species Concepts and Phylogenetic Theory* (New York: Columbia, 2000), pp. 17–29, cited at p. 17. This is a typical statement of the biological species concept. It is repeated word for word in several places: *Evolution and the Diversity of Life: Selected Essays* (Cambridge: Harvard, 1976), p. 518; *Populations, Species, and Evolution: An Abridgment of Animal Species and Evolution* (Cambridge: Harvard, 1970), p. 12; *Principles of Systematic Zoology* (New York: McGraw-Hill, 1969), p. 26. Still, the biological species concept has undergone change over Mayr's many active decades: see Peter Beurton, "Ernst Mayr through Time on the Biological Species Concept: A Conceptual Analysis," *Theory in Biosciences*, CXXI (2002): 81–98; and Mayr, "Comments by Ernst Mayr," *Theory in Biosciences*, CXXI (2002): 99–100.

[9] Cracraft, "Species Concepts and Speciation Analysis," in R.F. Johnston, ed., *Current Ornithology* (New York: Plenum, 1983), pp. 159–87, cited at p. 170.

[10] Cracraft, "Species Concepts in Systematics and Conservation Biology: An Ornithological Viewpoint," in M. Claridge, H. Dawah, and M. R. Wilson, eds., *Species: The Units of Biodiversity* (New York: Chapman and Hall, 1997), pp. 325–39, p. 331; Graham Martin, "Birds in Double Trouble," *Nature*, CCCLXXX (1996): 666–67.

Both concepts have pros and cons. Convenience, the politics of conservation, and other such issues hang in the balance. One salient issue between proponents of different concepts is the issue of whom systematists should serve. An alleged advantage of the biological species concept is that it is user friendly in the field. It takes a laboratory specialist to distinguish species according to the phylogenetic species concept. On the flip side, specialists can more readily convey detailed information to one another if lineages are divided into species according to the phylogenetic species concept. Differences between proponents of the phylogenetic species concept and the biological species concept hinge largely on whether systematists think the discipline should be practiced "as a science dominated by field work or by laboratory investigation."[11] And whether the discipline should be practiced with field work or laboratory work as the first priority is evidently not something that scientists can *discover* an answer to. No one concept seems to trump its competition in such a way that we could learn that it is the single proper answer to the nature of species.[12]

III. A DIVERSITY OF TREES

Just as there may be more than one natural arrangement of a given group of organisms into species, so there may be more than one natural arrangement of a given group of species into historical hierarchies. Consider the common situation in which a controversial lineage qualifies, according to the biological species concept, as a species divided into two subspecies, while for the phylogenetic species concept it counts as two different species. Of course, there is already a need for two different accounts of branching to represent the two options, if all species are to be depicted. One account will recognize two branches, the other just one. But it might be thought that that the problem stops here: that, while there may be subjectivity in determining how *many* species there are, there remains a single correct ordering of whatever species one acknowledges into groups that are united by common ancestry. The problem is not constrained in this way.

The subjectivity in counting species introduces subjectivity in the assignment of which organisms share a common ancestral species to the exclusion of other organisms. Assume that the two sides of the

---

[11] Martin, "Birds in Double Trouble," p. 667.

[12] For further discussion, see Joseph LaPorte, *Natural Kinds and Conceptual Change* (New York: Cambridge, 2004), chapter 3.

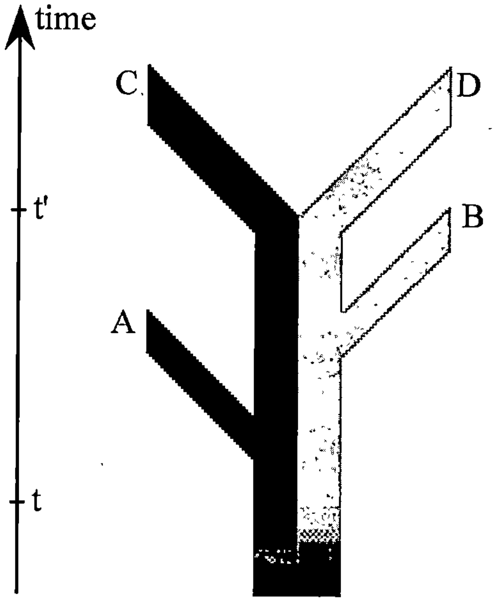


Figure 2: Between *t* and *t'*, the trunk is comprised of two borderline species.

trunk in figure 2 begin as one incontrovertible species. Later the species divides into two distinct lineages, represented in light and dark. There is a period of time, between *t* and *t'*, in which it is unclear whether the organisms on the trunk should be grouped into one species or two. There are marked differences between the sides, but whether there are enough differences or the right differences depends on whether the biological species concept or the phylogenetic species concept does the delimiting. Eventually, at time *t'*, the distinct sides separate into two species by any standard. (C) and (D) are the resulting daughter species. Before producing (C) or (D), each side of the trunk generates another daughter: species (A) or (B). These too are distinct species by any standard.[13]

Familial relationships between species (A)–(D) vary depending upon whether the trunk is comprised of one or two species. If the

---

[13] This sort of budding from a parental species that survives without much change is probably common: most speciation events probably start with small, isolated founder populations: Mayr, "A Defense of the Biological Species Concept," in Wheeler and Meier, eds., pp. 161–66, p. 164; *The Growth of Biological Thought* (Cambridge: Harvard, 1982), p. 229; Mayr and Peter Ashlock, *Principles of Systematic Zoology* (New York: McGraw-Hill, 1991, 2nd ed.), p. 225.

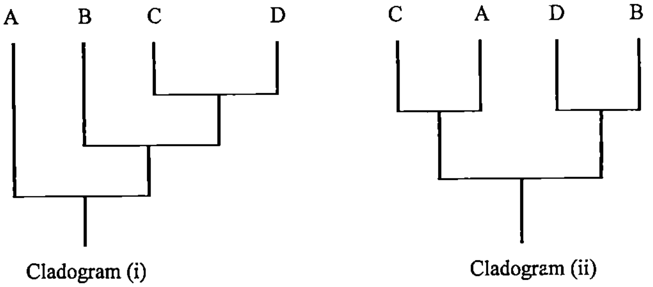


Figure 3: Two different accounts of the monophyletic groups. Cladograms nest groups within groups but give no further information. They do not indicate from which side of a parental lineage a branch issues, as figure 2 does: cladograms rotated at any node will express the same information.

division occurring at around *t* is only into subspecies, so that the trunk of the tree is still to be counted as one species, then the last two branches, (C) and (D), belong to an exclusive monophyletic group. A wider monophyletic group includes branch (B) and a still wider monophyletic group includes all branches. This interpretation, which favors the biological species concept, is represented by Cladogram (i).[14]

On the other hand, if the division prior to *t* is a division into separate species, in accordance with the phylogenetic species concept, then Cladogram (ii) is the proper one for representing the world. Now (C) and (D) are no longer sister species; rather, (C) is now a sister of (A), and (D) is now a sister of (B). There is no longer any monophyletic group containing just three of the branches. (B) is no longer a sister to a branch culminating in (C) and (D). And (A) is no longer a sister to the group culminating in all of the other branches.

If it is not a matter of discovery as to which species concept is correct, it likewise cannot be a matter of discovery as to which cladogram, (i) or (ii), is correct. There is no single, objectively correct branching diagram depicting the hierarchy of species.

If either the biological species concept or the phylogenetic species concept were the unique correct account of what a species is, contrary to what I have argued, then Cladogram (i) or (ii) might be discovered to be correct. So the species issue might be pressed. The most natural reason to hold out for a discovered resolution to the species problem is that one species concept may end up capturing the process or processes that produce and maintain the more or less homogeneous morphological groups of organisms to which biologists attach species names. The phylogenetic species concept might be dismissed on the grounds that it is not even a contender in the race to capture the right processes.[15] It attempts to identify groups that evolution has generated irrespective of the processes behind those groups' production. The biological species concept, on the other hand, might claim a measure of vindication if sexual recombination is sometimes responsible for cohesion. The right account might be a disjunctive one open to different kinds of processes.

---

[14] What is important is that the two-shaded trunk of the tree is counted as a single species at any one time (before *t'*). Whether the trunk must be split over time into three species, the endpoints set by branching events, is immaterial, because the order of branching for the species at the tips of cladograms (i) and (ii) is unaffected. I will explain below that for orthodox cladists, the trunk would have to be divided into three species, despite no intrinsic change in its members.

This reason to reject the phylogenetic species concept has scant claim to an objective discovery: no one could discover as an objective fact that Cracraft is just wrong in his use of species' names, on the grounds that what should really interest him is the processes producing the groups that he names rather than the groups defined by these processes.[16] Further, just as some theorists call attention to processes that give rise to broad groups, Cracraft could call attention to processes that give rise to the narrow groups that he recognizes, by reworking his statement. In this way, he could continue to split broad groups.

---

[15] Ridley, "The Cladistic Solution," pp. 3, 5–6, 8.

[16] *Objection*: the issue has been misconceived. The real issue is not whether Cracraft is wrong to name his narrower groups. One can apply names as one fancies. The real issue is whether Cracraft's use of 'species' breaks from traditional use. If there is a break from traditional use, then Cracraft errs in supposing that he is talking about *species*; he is talking about some other kind of group to which he has applied the name. *Reply*: traditional use does not settle in favor of any rival of the phylogenetic species concept. The tradition has long been shared by "lumpers" and "splitters." Both persuasions are reasonable. Over generations, workers of different persuasions have treated the same lineages differently: for examples, see Wiley and Richard L. Mayden, "The Evolutionary Species Concept," in Wheeler and Meier, eds., pp. 70–89, at p. 86; Jody Hey, *Genes, Categories, and Species: The Evolutionary and Cognitive Causes of the Species Problem* (New York: Oxford, 2001), pp. 19–20. The tradition has also long been characterized by different measures to delimit species, including patterns of resemblance, as Darwin indicates in a famous passage:

> It is really laughable to see what different ideas are prominent in various naturalists' minds, when they speak of "species"; in some, resemblance is everything and descent of little weight—in some, resemblance seems to go for nothing, and Creation the reigning idea—in some, sterility an unfailing test, with others it is not worth a farthing. It all comes, I believe, from trying to define the indefinable (*The Life and Letters of Charles Darwin, Including an Autobiographical Chapter*, Volume 2, F. Darwin, ed. (London: John Murray, 1887), p. 88).

Even the tradition of cladism fails to settle upon a single species concept. Hennig dodges the issue: see section VI. Hennig's most stalwart followers fail to agree on a single concept: cf. D.J. Kornet, "Permanent Splits as Speciation Events: A Formal Reconstruction of the Internodal Species Concept," *Journal of Theoretical Biology*, CLXIV (1993): 407–35; Rudolf Meier and Rainer Willmann, "The Hennigian Species Concept," in Wheeler and Meier, eds., pp. 30–43; "A Defense of the Hennigian Species Concept," in Wheeler and Meier, eds., pp. 167–78; Ridley, "The Cladistic Solution."

The best way to demonstrate the robustness of the problem may be to expand the discussion. The *ecological species concept*, like the biological species concept, identifies species as groups united by a certain kind of process. Just as sexual processes do, ecological processes sometimes produce groups that are relatively homogeneous, genetically and phenotypically. The organisms thus grouped include not only plants, oaks being the most famous example,[17] but also animals, such as certain species of North American snakes or Darwin finches, for which selection in the environment evidently eliminates hybrids between distinct forms:[18] hence, the recognition of ecologically delimited species.[19]

Sometimes, both ecological processes and sexual processes operate to shape the same organisms into groups. For that reason, a large lineage like that represented by the trunk in figure 2 could easily be divided into subgroups, the big group qualifying as a species according to the biological species concept and the narrower subgroups qualifying as separate species according to the ecological species concept, or vice versa. There might be no clear, objective answer as to whether one species concept prevails over the other in characterizing the lineages: "the nature of the evolutionary entity could be inherently ambiguous," as a group of biologists observes in this connection.[20]

IV. THE MANY TREES FOR ANY SPECIES CONCEPT

In criticizing the proposal that there is one objective hierarchy of species, I have appealed to competing species concepts. But the source of the problem lies deeper. Suppose, contrary to what seems reasonable, that one species concept or another like the biological species concept or the phylogenetic species concept were somehow to have an objective claim to be the sole accurate species concept. Still, there would be room for doubtful lineages like that shown to give rise to cladograms (i) and (ii).

---

[17] Alan Templeton, "The Meaning of Species and Speciation: A Genetic Perspective," in Marc Ereshefsky, ed., *The Units of Evolution* (Cambridge: MIT, 1992), pp. 159–83, p. 174; L. Van Valen, "Ecological Species, Multispecies, and Oaks," in Ereshefsky, pp. 69–77, pp. 72–73.

[18] Werner Kunz, "When Is a Parasite Species a Species?," *Trends in Parasitology*, XVIII (2002): 121–24.

[19] Van Valen, "Ecological Species"; L. Andersson, "The Driving Force: Species Concepts and Ecology," *Taxon*, XXXIX (1990): 375–82; see also Ridley, "The Cladistic Solution."

[20] Jody Hey, Robin S. Waples, Michael L. Arnold, Roger K. Butlin, and Richard G. Harrison, "Understanding and Confronting Species Uncertainty in Biology and Conservation," *Trends in Ecology and Evolution*, XVIII (2003): 597–603, p. 599.

Take the biological species concept. It may seem that,

> The intra-breeding criterion yields only *one* set of groups of living creatures, and the determination of which creatures to assign to the same group and which to different groups is no longer in any sense a matter of choice but only of discovery.[21]

But in fact the distinctness generated comes in degrees. Systematists find it necessary to recognize that distinct species can cross to some degree, despite substantial reproductive isolation. Otherwise, species would be far larger and more heterogeneous than it would be natural to allow. Unfortunately, there would seem to be no fact of the matter as to precisely how much potential for hybridization may be tolerated as species are delimited.[22]

Even if systematists were willing to go to extreme measures by lumping otherwise good species that can hybridize into a single species, they could not in this way circumvent the problem of indistinctness.[23] There is no distinct line at which absolute reproductive isolation can be said to have been achieved: it may be that certain environmental circumstances, for example, preclude any interbreeding even though a change in environment could alter this.[24] This and other troubles render the notion *reproductive isolation* unclear and incapable of black and white application.

---

[21] Sybil Wolfram, *Philosophical Logic: An Introduction* (New York: Routledge, 1989), p. 239.

[22] A limited amount of hybridization is common among recognized species of both plants and animals: for discussion of examples, see C. J. Hearn, "The Evolution of Citrus Species: Methods to Develop New Sweet Orange Cultivars," *Proceedings of the Annual Meeting of the Florida State Horticultural Society*, CVII (1994): 1–3, p. 2; Daniel Potter, Fangyou Gao, Scott Baggett, James R. McKenna, and Gale H. McGranahan, "Defining the Sources of Paradox: DNA Sequence Markers for North American Walnut (*Juglans* L.) Species and Hybrids," *Scientia Horticulturae*, XCIV (2002): 157–70; Douglas J. Futuyma, *Evolutionary Biology* (Sunderland, MA: Sinauer, 1998, 3rd ed.), p. 448. The problem of drawing a line at some point has long been recognized: "the infertility of species when crossed graduates away so insensibly that the two most experienced observers who ever lived have come to diametrically opposite results when experimentising on the same forms," Darwin writes (*Charles Darwin's Natural Selection: Being the Second Part of His Big Species Book Written From 1856 to 1858*), R.C. Stauffer, ed. (New York: Cambridge, 1975), p. 165). The same observation has been made more recently by Theodosius Dobzhansky, *Genetics of the Evolutionary Process* (New York: Columbia, 1970), p. 359; Hey, *Genes, Categories, and Species*, pp. 20–21; Johnson, "Rainbow's End," p. 231; Mayr, *The Growth of Biological Thought*, p. 282; B. Mishler and M. Donoghue, "Species Concepts: A Case for Pluralism," *Systematic Zoology*, XXXI (1982): 491–503, p. 495; Templeton, "The Meaning of Species and Speciation," p. 165.

[23] Cf. Meier and Willmann, "A Defense." This suggestion raises the possibility that the term 'species' be given by stipulation a new use that would make tree construction a repeatable enterprise. Occasions of discord could be *reduced* in this way: even stipulating the biological species concept over the phylogenetic species concept is an improvement. But there is little value in going to drastic measures in this direction to describe a kind of group that, however quirky or biologically meaningless, allows for repeatable ordering. Nor would any algorithm generating an ordering of quirky groups show that there is a repeatable account of the hierarchy of *species*; at best, such an algorithm would show that there is a repeatable account of the hierarchy of some new kind of group deserving a new name. Observe that just as 'species'

So even if we restrict our attention to the biological species concept, we are bound to find cases like that depicted in figure 2, in which there is no definitive answer, even in principle, as to whether there are two species present or just one species divided into two subspecies. Cursory reflection on the phylogenetic species concept and the ecological species concept indicates that the application of these other species concepts cannot be any more cut and dried than the application of the biological species concept. Still other species concepts fare the same.[25] Evolution assures that they do. As the great G. G. Simpson observes, "To insist on an absolute objective criterion would be to deny the facts of life, especially the inescapable fact of evolution."[26]

V. WHAT IS OBJECTIVE AND WHAT IS NOT?

Not all groups of species permit more than one possible arrangement by order of descent. When does reality uphold just one account? In the present section, I try to clarify this.

When all but a few species are ignored, historical relationships can easily be closed to interpretation. Consider the relationships between any two species of gecko and the lizard-like tuatara species *Sphenodon punctatus*, whose ancestral lineages have been separated from those of all other living reptiles for hundreds of millions of years. Because the tuatara line and the gecko line have been separate for so long, each has myriad common ancestral species that are not shared by the other, *no matter how species are delimited*. Any arrangement according to which the tuatara and one gecko species belong to a monophyletic group that excludes another gecko species is just wrong. Gaps between species produce lots of relationships like these. Wolves and foxes

---

might be given an artificial, stipulated use that approximates to some extent ordinary use in order to allow for repeatable groupings, so might 'similarity': tellingly, cladists find this trick for achieving repeatability to be arbitrary and to preclude the kind of objectivity in ordering that cladism is supposed to enjoy (Ridley, "Can Classification Do Without Evolution?," p. 199).

[24] LaPorte, *Natural Kinds and Conceptual Change*, chapter 2; Philip Kitcher, "Some Puzzles about Species," in Michael Ruse, ed., *What the Philosophy of Biology Is: Essays Dedicated to David Hull* (Boston: Kluwer, 1989), pp. 183–208, pp. 197–98; Ridley, *Evolution*, p. 358.

[25] I survey some other concepts in chapter 3 of *Natural Kinds and Conceptual Change*.

[26] *Principles of Animal Taxonomy* (New York: Columbia, 1961), p. 152.

separated around five million years ago, but not before they shared tens of millions of years of ancestral species outside the line leading to bears. Humans and chimps share a common ancestor not shared by gorillas. Oranges and lemons share a common ancestor not shared by watermelons. The lesson of this paper does nothing to upset the objectivity of many hierarchies like these.

There is much to be said for objectivity, then. When we prune the tree to a few branches, no two of which share exclusively a closest common ancestor that lies close to the origin of any other branch, then we obtain an ordering that admits of no legitimate dissent, in view of the empirical facts.

But if there is much to be said for objectivity, there is much to be said against it. Relationships are thrown into turmoil when there is no whittling the tree down, or when groups left on the tree have more closely related ancestral species. These relationships can be of great interest, as I will illustrate with an example close to home. A massive amount of effort from multiple disciplines has been expended in the effort to discern the relationships between our own species and our closest relatives in the genus *Homo*.

The genus *Homo* is divided into varying numbers of species, depending on whether differences in specimens at various geographical locations are seen to represent a departure significant enough to warrant the conferral of a distinct species name. Some theorists lump African and Asian specimens of our best-known ancestor all into the same species, which is then called "*Homo erectus*." Others see the differences between African and Asian lineages to be significant enough to warrant different specific designations: these theorists group African specimens into "*Homo ergaster*," reserving '*Homo erectus*' for the Asian lineage. There is no particular reason to suppose that there is an objective answer as to whether there are two species or one here: potential interbreeding is difficult to determine, but it is certainly reasonable to suppose that by one species concept, say the biological species concept, there may be one species divided into two subspecies, while for another concept, say the phylogenetic species concept, there are enough differences between the strains to merit two specific names. As I have indicated, such situations are not rare.

Homo ergaster (or, depending on your perspective, African *Homo erectus*) and the Asian strain of *Homo erectus* are thought to have diverged over a million years ago. Hundreds of thousand of years later, *Homo ergaster* (or African *Homo erectus*) seems to have given rise to the Neandertals, *Homo neanderthalensis*, and then our own species, *Homo sapiens*. Meanwhile, the Asian strain of *Homo erectus* seems to have given rise to the "hobbit" species *Homo floresiensis* that shared the

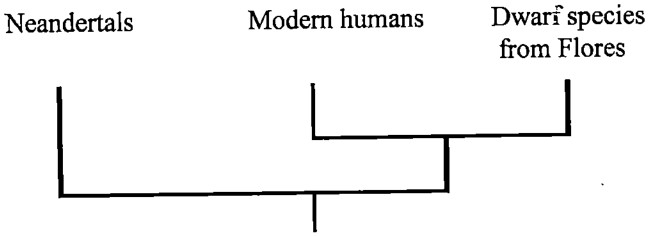


Figure 4

planet with us until only a few thousand years ago. The remarkable discovery of this new species has made headlines in recent months.

No one is quite sure about the precise dates at which these more recent species emerged, or the order in time. But for the sake of illustration, I will suppose that after the split between *Homo ergaster* (or African *Homo erectus*) and the Asian strain of *Homo erectus,* the next split was that which resulted in *Homo neanderthalensis* (from the African line). Sometime after the emergence of *Homo neanderthalensis,* there emerged, in one order or another, both our own species, *Homo sapiens* (from the African line), and *Homo floresiensis* (from the Asian line). This is the correct order of events, for all anyone knows.[27]

Now let us suppose that the strain that I have been calling "*Homo ergaster or* African *Homo erectus*" is to be counted as conspecific with Asian *Homo erectus,* so that the two comprise the single species *Homo erectus* (between branching events: see note 14). In that event, *Homo neanderthalensis* is the species first to separate from others in the group: it branches off from a species that will go on to generate both *Homo sapiens* and *Homo floresiensis. Homo sapiens* and *Homo floresiensis* are more closely related to each other than either is to *Homo neanderthalensis.* This is represented in figure 4 above.

If, on the other hand, the strain that I have been calling "*Homo*

---

[27] *Homo neanderthalensis* is typically thought to be around 400,000 years old and *Homo sapiens* around 200,000 years old. *Homo floresiensis* may be as young as 70,000 years, as preliminary dating for the species indicates, but it could also be much older, perhaps 800,000 years old: Rex Dalton, "Little Lady of Flores Forces Rethink of Human Evolution," *Nature,* CDXXXI (2004): 1029; see also Marta Mirazón Lahr and Robert Foley, "Palaeoanthropology: Human Evolution Writ Small," *Nature,* CDXXXI (2004): 1043–44, at p. 1043. If the earlier date is accurate, *Homo floresiensis* may still be younger than *Homo neanderthalensis,* whose roots may lie further back in Africa than the date of origin just mentioned: Chris Stringer, "Human Evolution: Out of Ethiopia," *Nature,* CDXXIII (2003): 692–95, p. 693.

Further information might, of course, reveal that the order in which the three species in question emerged is not as I have assumed. But even if my example becomes outdated for this reason, minor alterations could probably render it historically accurate: some other lineage could be substituted in place of *Homo floresiensis* to the same effect. As one of the team leaders responsible for finding *Homo floresiensis*

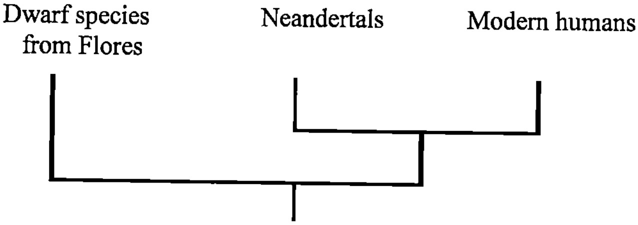


Figure 5

ergaster or African *Homo erectus*" is *not* to be counted as conspecific with Asian *Homo erectus*, so that the two comprise the pair of species *Homo ergaster* and *Homo erectus*, then a different cladogram is needed. *Homo floresiensis* is a daughter of the Asian species *Homo erectus*, while *Homo sapiens* and *Homo neanderthalensis* are daughter species of *Homo ergaster*. *Homo sapiens* and *Homo neanderthalensis* are sister species, sharing a more recent common ancestor than either shares with *Homo floresiensis*. *Homo floresiensis* is the outsider. This is represented in figure 5.[28]

The number of distinct trees that might be natural and legitimate

---

indicates, *Homo floresiensis* was probably one of many species to arise during human history from populations isolated on islands: "I think we're going to have a plethora of new human species showing up," Michael Morwood says (quoted in Michael Lemonick, "Hobbits of the South Pacific," *Time* (November 8, 2004): 50–52, at p. 52); the same claim is made by P. Brown, T. Sutikna, M. J. Morwood, R. P. Soejono, Jatmiko, E. Wayhu Saptomo, Rokus Awe Due, "A New Small-Bodied Hominin from the Late Pleistocene of Flores, Indonesia," *Nature*, CDXXXI (2004): 1055–61, at p. 1061.

[28] Figures 2 and 3 may also illustrate the different possible scenarios that I have depicted for *Homo*, provided (A) represents *Homo neanderthalensis*, (B) represents *Homo floresiensis*, (C) represents *Homo sapiens*, and (D) represents the rump of *Homo erectus* left in Asia after the speciation event that produced *Homo floresiensis*. On this assignment, the trunk represents *Homo ergaster* or African *Homo erectus* both at its base and on the left in black. The lighter, right side of the trunk, after the division into distinct shades, represents the Asian *Homo erectus*.

It would be possible to object that because *Homo ergaster* is the closest common ancestor of *Homo sapiens* and *Homo neanderthalensis*, and it generates *Homo erectus*, which generates *Homo floresiensis*, *Homo floresiensis* shares the closest common ancestor of *Homo sapiens* and *Homo neanderthalensis* after all: the closest common ancestor of any two of all three species is *Homo ergaster*. This objection will not go far. For those who would be tempted by it (orthodox cladists would not, for reasons indicated in note 14), I would preclude this interpretation by introducing another species between *Homo ergaster* and the sisters *Homo sapiens* and *Homo neanderthalensis*, which is then shared by these two sisters but not by *Homo floresiensis*. Indeed, species splitters, who resist lumping *Homo ergaster* with *Homo erectus*, do insert additional species, which are all lumped into "*Homo erectus*" by lumpers: see, for example, Stringer, "Human Evolution." For the sake of simplicity I have refrained from bringing to the discussion such species as *Homo heidelbergensis*, which may be inserted between *Homo ergaster* and the sisters *Homo sapiens* and *Homo neanderthalensis*.

multiplies when we recognize a more extended family. The species giving rise to early members of *Homo erectus* (or *Homo ergaster*) may have been *Homo habilis.*[29] But various authorities split *Homo habilis,* too, into two different species.[30] And so it goes.

I have indicated only part of the legitimate variation in tree construction even for the lineages depicted. Some authorities lump Neandertals into our own species, counting them members of *Homo sapiens.*[31] These authorities see Neandertals as a distinct subspecies of modern humans. It is possible that missing information could, if provided, settle the disagreement: for example, definitive information that there were genetic barriers to interbreeding could settle the issue.[32] However, matters could as well be more complicated than that: there is some evidence of *limited* interbreeding where the lineages coexisted.[33] It might be that no amount of information could settle the issue to the satisfaction of authorities in general, since authorities who agree on the empirical facts might simply find it useful to delimit lineages differently, as do authorities on birds who disagree about the number of existing species.

The more we find about species and their close relatives, the more we are bound to find that different trees are permissible. Groups of three or more closely related species are obviously more likely to have closely related nearest ancestral species than groups whose ancestors meet further back. They are more likely to make trouble. The hierarchy is straightforward for lions, bears, and canines; three canine species are another story. The hierarchy is straightforward for pines, watermelons, and citruses; three citrus species are another story.

VI. SYSTEMATICS AGAIN

The observation that there are different legitimate trees has practical interest. And it has obvious bearing on the work of those who are trying to obtain a picture of the whole panorama of life. Both points can be seen in implications for the work of conservationists, as a cursory reflection will indicate. Both points can also be seen in implications for the practice of systematics, as a cursory reflection will *not* indicate. I return to the connection to systematics in the present section.

---

[29] Barbara King, *Biological Anthropology: An Evolutionary Perspective, Part I* (Chantilly, VA: Teaching Co., 2002), pp. 164, 170; Monroe W. Strickberger, *Evolution* (Sudbury, MA: Jones and Bartlett, 2000, 3rd ed.), p. 483.

[30] King, *Biological Anthropology, Part I,* p. 159.

[31] An example is Strickberger, *Evolution,* p. 483.

[32] King, *Biological Anthropology: An Evolutionary Perspective, Part II* (Chantilly, VA: Teaching Co., 2002), p. 17.

[33] Stringer, "The Evolution of Modern Humans: Where Are We Now?," *General Anthropology,* VII (2001): 1–5, at p. 3.

In the practice of systematics, as I have observed, cladists tend to trumpet the need for one legitimate, objective, unambiguous ordering of living groups. Cladists also insist upon cladism's unique capacity to fulfill that need. There is some truth to the claim that the order of evolutionary branching is objective: it is objective if you ignore troublesome species. But that is to ignore a lot of texture in the history of species. It is no wonder, then, that cladists often advertise that one can in principle capture *all* texture down to the level of branching species in a unique, legitimate rendering of the order of species: the objectivity is supposed to start at the level of distinguishing sister species. Hennig is supposed to show how species should be delimited in time in order to allow for objectivity in the phylogenetic hierarchy of species.[34]

According to Hennig, every species originates with a speciation event and goes into extinction with a speciation event.[35] A speciation event is a branching forth of a new species. Each species begins with a forking event and ends when it itself forks into separate species (see figure 6).[36] When does speciation occur? Is it, for example, when a species divides into groups that are no longer interfertile that the species goes extinct and new species are born? Hennig has little to say about this.

Hennigian constraints on the recognition of species are controversial. Because branching is necessary for speciation, Hennigians cannot recognize *anagenetic* speciation. Thus, species D in figure 6 cannot be divided into more than one species (D1, D2, and so on) no matter how much the D-lineage changes in genetic constitution, breeding habits, and so on, between the numbered segments and no matter how long the time period from each numbered segment to the next; only one species can be recognized because there are no branching events between the segments.

Branching is also sufficient for speciation. This can be awkward if there is little change in a branching lineage: thus, in figure 6, species *A* goes extinct upon the arrival of *C*, even though the lineage including *A* and *B* does not evolve at all.

---

[34] Kornet, "Permanent Splits"; Meier and Willman, "The Hennigian Species Concept"; "A Defense," p. 167; Ridley, "The Cladistic Solution," pp. 2–4.

[35] Hennig, *Phylogenetic Systematics*, p. 66.

[36] For the sake of recognizing hybrid species and species that leave no descendants, Hennig's definition must be slightly modified.

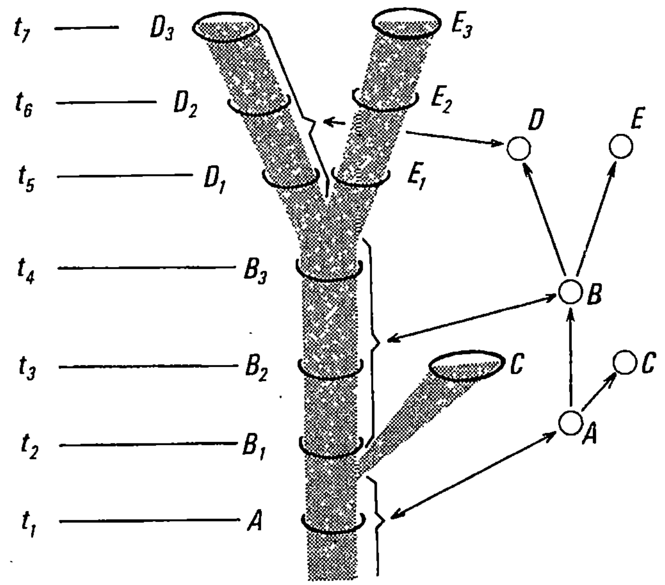


Figure 6: Cladistic species.[37]

In actual practice, authorities are not apt to split species like *Homo erectus* in time wherever there is branching; no one names a new species where *Homo floresiensis* branches off. And authorities do recognize new species without knowledge of branching.[38]

No doubt there are epistemic reasons for recognizing species' survival through speciation or for recognizing speciation in the absence of branching. One might know that two preserved specimens belong in a common lineage without having much information about whether any speciation events have occurred in the time between the two specimens. Then it might seem to make sense to place the two specimens in a common species, even though there is no evidence one way or the other whether the specimens could fulfill Hennig's criteria for conspecificity. In the same way, one might know that substantial differences distinguish two preserved specimens. In that case, it might seem sensible to place the organisms into different species even if there is not much information about whether any speciation event separates the two specimens or whether, instead, the changes between the earlier and later forms came about by way of evolution in a single unbranching line.

---

[37] From W. Hennig's *Phylogenetic Systematics*. Copyright 1966, 1979 by the Board of Trustees of the University of Illinois. Used with permission of the University of Illinois Press.

[38] See Strickberger, *Evolution*, p. 483; Lahr and Foley, "Palaeoanthropology"; Stringer, "Human Evolution."

Hennigian species are sometimes rejected for reasons having nothing to do with our limited state of knowledge: Mayr complains that Hennigian constraints produce "biological nonsense."[39] A look at actual classification indicates that other practicing systematists agree: even where records about branching are very complete, species are often not taken to be delimited in time by branching events.[40]

There are, then, initially plausible considerations *against* Hennigian constraints on the delimiting of species. There are also initially plausible considerations in *favor* of Hennigian constraints on the delimiting of species, as I have indicated: orthodox cladists insist that only the Hennigian species concept preserves the objectivity of the resulting classification. A complete classification of species and their place in the hierarchy of life ceases to be objective if systematists are allowed to make judgment calls about when enough change marks a lineage to legitimate a decision to recognize that a new species has emerged.

A lesson of the foregoing sections is that initially plausible considerations in favor of Hennigian constraints are overstated. The Hennigian species concept does not preserve any single, unambiguous account of species and their relationships to one another. There is no such account, not even in principle.

If Hennigians are to make a decent case for a general adoption of their proposed practice, it will have to be better than the familiar fare. Without such an argument, there is little theoretical barrier to the common practice of delimiting the endpoints of species in other ways.

VII. CONCLUSION

We may whittle a tree of life, paring troublesome branches, in order to portray an ordering that admits of no legitimate dissent. But the history of life can sustain many legitimate arrangements of the same branches. This observation affects relationships of much interest. It undermines a thoroughgoing cladistic systematics.

JOSEPH LAPORTE

Hope College

---

[39] "A Defense," p. 164.

[40] Meier and Willmann acknowledge this, citing species of freshwater gastropods: "A Critique from the Hennigian Species Concept Perspective," in Wheeler and Meier, eds., pp. 101–18, p. 105; see also Wiley and Mayden, "A Critique from the Evolutionary Species Concept Perspective," in Wheeler and Meier, eds., pp. 146–58, at p. 157.

COMMENTS AND CRITICISM

A NOTE ON GOODMAN'S PROBLEM*

In "Goodman's Problem and Scientific Methodology,"[1] Peter Godfrey-Smith argues that the troublesome hypotheses Nelson Goodman introduced can be eliminated by existing methodological principles of scientific inference. It seems to me, however, that although Godfrey-Smith explores some interesting features of the New Riddle, his account is no more of a solution than others published at earlier times *t*.

Godfrey-Smith divides "bad" projections into two broad classes: those eliminated by rules of data analysis and those eliminated by practices of inference "about the structures and mechanisms responsible for the clustering of properties in a kind" (587). I will treat each in turn. A few caveats before I begin. First, an adequate solution to the problem of projection should not depend on the specific properties of color and time associated with the "grue" example. Goodman's original challenge to Rudolf Carnap depends on patterns found in the evidence base that conflict when projected.[2] The riddle can also be presented as a dilemma about curve fitting, that makes no reference to facts about observation and time.[3] Indeed, Israel Scheffler shows how any predicate that characterizes the evidence can serve, on an assumption of conflict, to provoke the problem.[4]

THE DATA ANALYSIS SOLUTION

Godfrey-Smith argues that his data analyses examples, themselves, are best split into two kinds: (a) those like "All emeralds are grue" and (b) those like "All emeroses are green." Different methodological principles may be employed in eliminating each type of troublesome hypothesis. "Grue" hypotheses are ruled out, because they introduce confounding variables; "emerose" hypotheses can be discarded, because the evidence cited is biased. Confounding variables and biased

---

* Thanks to Is Scheffler for some helpful suggestions.

[1] This JOURNAL, C, 11 (November 2003): 573–90

[2] "Infirmities of Confirmation-Theory," *Philosophy and Phenomenological Research*, VIII (1947): 133–48

[3] Carl Hempel, "Inductive Inconsistencies," in *Aspects of Scientific Explanation* (New York: Free Press, 1964), pp. 53–79. See also James Hullett and Robert Schwartz, "Grue: Some Remarks," this JOURNAL, LXIV, 9 (May 11, 1967): 259–71.

[4] *Anatomy of Inquiry* (New York: Knopf 1963), pp. 295–310.

0022-362X/05/0207/375–79

samples establish "spurious associations" that give rise to the inductive paradoxes. There is, of course, no need to deny that the New Riddle can be seen to result from spurious associations. The problem is to distinguish associations that are spurious and from those that are not.

Throughout his paper Godfrey-Smith makes clear that he is operating with(in) Goodman's own notion of "grue," that is, grue=df (green & examined before *t*) ∨ (blue & not examined before *t*). On this definition, a grue emerald is *not* one that changes its color at *t*. Now, Godfrey-Smith claims that the confounding nature of Goodman's predicate 'grue' stems from its semantics, a peculiarity in its application. If emerald *a* is green, the predicate 'green' applies to it whether the emerald has been observed {*O*} or has not been observed {−*O*}. The semantics of 'grue', he argues, are different. If emerald *a* is grue and has been observed, it is green, whereas if it has not yet been observed, it would have to be blue for 'grue' to apply. The property or additional variable, "being observed," plays a role in determining which items fall under the predicate 'grue'. But we know/believe that acts of observation do not alter an object's color. Thus the semantics of 'grue' introduces a variable that confounds. The projection of 'grue' depends on *spurious* associations that violate accepted methodological principles, in this case the principle concerning the irrelevance of processes of observation to color predication (580).

This solution to the New Riddle, however, faces several difficulties:

(1) As it stands, Godfrey-Smith's criterion for hypothesis elimination is not readily applicable to cases where the paradox is stated in terms of conflicting evidence patterns or curve fitting.

(2) Success in handling these cases, would still leave extensive work undone. Any predicate meeting Scheffler's conditions can serve to generate the puzzle. A comprehensive resolution of Goodman's problem should seek to eliminate them all, while not excluding "good" hypotheses. Consider, for example, Scheffler's version of the riddle that depends on an assignment of names to objects in the evidence class (*op. cit.*). Scheffler defines 'grue' as: grue=df (is identical to *a*, *b*, or *c* & green) ∨ (is not identical to *a*, *b*, or *c* & blue). This predicate 'grue' can fail to apply to a green emerald in the evidence class only on the assumption that the emerald is not self identical. So somewhat different considerations or additional assumptions will likely be involved in showing that Scheffler's 'grue' implicates an extraneous property in the way "being observed" is said to affect a predicate's application. And remember, in the context of Goodman's problem, elaborations or adjustments to this methodological solution that de-

pend on counterfactuals, lawlikeness, essential properties, necessities, and similar notions are of dubious worth.[5]

(3) Godfrey-Smith's solution to the riddle, moreover, runs into trouble with Goodman's original paradigm case of 'grue' versus 'green'. To begin, suppose we introduce the predicate '*P*', P=df (green & round) $\vee$ (blue & square). The situation with respect to the application of '*P*' is analogous with that of 'grue'. Were emerald *a* green and round, it is *P*, but were *a* green and square, '*P*' would not apply. Is '*P*', too, then be eliminated? Right off, it is not obvious that it is very peculiar *semantically*. And even if it is, there is no strong reason to suppose that *formal* properties or principles of semantics can provide a basis for distinguishing "normal" from "odd" predicates in general. In any case, merely convicting 'grue' of semantic strangeness should not be all that crucial to the matter at hand. According to Godfrey-Smith, the real methodological problem with 'grue' is that its semantics invites spurious associations. If a predicate is peculiar, but does not also *confound*, there should be no pressing need to eliminate it.

Is it true that 'grue' does form a spurious association with '*O*' and 'green' does not? Suppose 'green' is defined as: green=df (grue & O) $\vee$ (bleen & $-O$). The application of 'green', then, also depends on *O*. If emerald *a* is *O*, it is grue, and if *a* is $-O$, it is bleen. Thus projecting 'green' would seem to violate the same methodological principle concerning observational dependence and color hitherto brought in to eliminate projection of the "grue"-hypothesis.

An initially enticing response to this *symmetry* argument, perhaps, is to claim that grue and bleen are not "really" colors. There is no color difference involved in the unconditional application of 'green' as there is with 'grue' and 'bleen'.[6] But why accept this intuition? It would not advance matters to defend the intuition on grounds that 'green' is a semantically good predicate and 'grue' and 'bleen' are bad. Support for the intuition, seems to rest more on various established practices of color categorization. We assume, for example, that blue and green are different colors, because they look different; they are visually distinctive. Things that are (green & O) are not appropriately similar to those that are (blue & $-O$), so they differ in color. Things that are (grue & O) do match those that are (bleen & $-O$), so they are of the same or appropriately similar color.

---

[5] In several places Godfrey-Smith acknowledges the shortcomings and unsatisfactory nature of several solutions that do not take these constraints into account. It should be clear in what follows that my main qualms with Godfrey Smith's paper are due neither to his subdividing the puzzle into distinct cases, nor to the fact that he does not handle all possible versions of the New Riddle. I question the advance his analysis offers in treating the cases he does attempt to settle.

[6] Note, though, one can introduce a predicate 'color*' that differs from 'color' in ways comparable to and as paradox generating as the "green"/"grue" pair. Application of the predicate 'green' will reflect an observational dependence on color* much as 'grue' does with respect to color. I will not explore explicitly the further difficulties this possibility raises for Godfrey-Smith's proposal.

Were these practices of color term usage to reflect habits of categorization (that is, entrenchment), Goodman may have no complaint with this analysis. Still, such tests of property identity or similarity are quite parochial. These notions of "likeness" and "matching" can *not* be meaningfully employed with respect to a wide range of everyday and scientific properties, nor to many of the predicates found in the alternative versions of the New Riddle mentioned at the start. Extending Godfrey-Smith's "grue solution" requires a *general* characterization of REAL property identity/similarity, one that is independent of customary or scientifically accepted ways of grouping objects into kinds.[7]

Although Godfrey-Smith treats the "grue" hypotheses and the "emerose" hypotheses separately, his solution to the emerose puzzle encounters difficulties akin to those involving "grue."[8] A sample of emeroses examined before t is said to be biased, since it contains only emeralds. Roses are in the projective class, but no rose examined before t is an emerose. Unfortunately, the situation is symmetrical if 'emerald' is defined as: emerald =df (emerose & O) $\vee$ (rosem & $-O$). Before t, the evidence class for "All emeralds are green" consists solely of emeroses and no rosems, yet rosems are in the projection class of the hypothesis. Needless to say, something is peculiar about the predicates 'emerose' and 'rosem'. But to say only this is to restate the New Riddle, not solve it.

CAUSES AND MECHANISMS

Godfrey-Smith points out that his confounding variables and biased sample solutions will not eliminate numerous "bad" inductive hypotheses. He does not see this as a defect. In fact, a major claim of his paper is that there are a variety of inductive paradoxes, and they require a variety of distinct solutions. In section IV, Godfrey-Smith sketches several additional methodological considerations that may be brought in to do the job. Goodman, surely, would agree with much of the overall picture. For him, too, grue is "an odd-looking property in relation to our background knowledge" (585). Goodman also maintains that entrenchment is not the sole methodological consideration influencing inductive policies. Simplicity and factors of conservativeness, for example, do play roles—albeit on Goodman's account, these pragmatic features of sound science have links with entrenchment and are dependent on the predicate scheme in play.

---

[7] I emphasize "real" as a reminder that if emerald a is grue, a has and keeps this property unconditionally or tenselessly, just as it does its greenness.

[8] For purposes of his analysis of a solution to the New Riddle traced to Jackson-Pargetter-Bigelow, Godfrey-Smith introduces the predicate emerose2 [emerose2 =df (emerald) $\vee$ (rose & $-O$)], in addition to his more Goodman-like emerose1 [emerose1=df (emerald & O) $\vee$ (rose & $-O$)]. This difference is not particularly germane to my point. Moreover, Godfrey-Smith has his own reasons for challenging the Jackson-Pargetter-Bigelow approach.

Godfrey-Smith admits that his attempt in section IV to extend his methodological solution involves substantive assumptions. And herein lies a weakness. The sticking point, though, is not that the assumptions are substantive. The problem is that they tend to be question begging. The scientific practices and background knowledge Godfrey-Smith cites themselves depend on prior habits of categorization and the attendant distinctions between projectible and nonprojectible hypotheses they inform. Godfrey-Smith's critical reliance in section IV on notions of "natural kinds," "the right class for analysis," "mere *sub*populations," "nomic properties," and other ideas "about the clustering of properties in a kind" make this dependence apparent.

ROBERT SCHWARTZ

University of Wisconsin–Milwaukee

NEW BOOKS: TRANSLATIONS

MALABOU, CATHERINE, and JACQUES DERRIDA: *Counterpath: Traveling with Jacques Derrida.* DAVID WILLS, trans. Stanford: University Press, 2004. xvi, 334 p. Cloth $50.00, paper $19.95.

MICHELSTAEDTER, CARLO: *Persuasion and Rhetoric.* RUSSELL SCOTT VALENTINO, CINZIA SARTINI BLUM, and DAVID J. DEPEW, trans. New Haven: Yale, 2004. xxvii, 178 p. Cloth $32.50.

OPHIR, ADI: *The Order of Evils: Toward an Ontology of Morals.* RELA MAZALI and HAVI CAREL, trans. Cambridge: MIT, 2005. 699 p. Cloth $38.50.

PASCAL, BLAISE: *Pensées.* ROGER ARIEW, ed. and trans. Indianapolis: Hackett, 2005. xix, 328 p. Cloth $34.95, paper $11.95.

PERNIOLA, MARIO: *Sex Appeal of the Inorganic.* MASSIMO VERDICCHIO, trans. New York: Continuum, 2004. 160 p. Cloth $90.00, paper $27.95.

PHILOPONUS: *Against Proclus' On the Eternity of the World 1-5.* MICHAEL SHARE, trans. Ithaca, NY: Cornell, 2005. ix, 154 p. Cloth $72.50; *Against Proclus' On the Eternity of the World 6-8.* MICHAEL SHARE, trans. Ithaca, NY: Cornell University Press, 2005. viii, 200 p. Cloth $75.00; *On Aristotle's On the Soul 2.1-6.* WILLIAM CHARLTON, trans. Ithaca, NY: Cornell, 2005. viii, 184 p. Cloth $79.95; *On Aristotle's On the Soul 2.7-12.* WILLIAM CHARLTON, trans. Ithaca, NY: Cornell, 2005. x, 213 p. Cloth $79.95.

PLATO: *Republic.* C.D.C. REEVE, trans. Indianapolis: Hackett, 2004. xxxiv, 358 p. Cloth $37.95, paper $9.95.

RANCIÈRE, JACQUES: *The Flesh of Words: The Politics of Writing.* CHARLOTTE MANDELL, trans. Stanford: University Press, 2004. ix, 170 p. Cloth $50.00, paper $19.95.

THEUNISSEN, MICHAEL: *Kierkegaard's Concept of Despair.* BARBARA HARSHAV and HELMUT ILLBRUCK, trans. Princeton: University Press, 2005. xi, 169 p. Cloth $35.00.

TICHY, PAVEL: *Pavel Tichy's Collected Papers in Logic and Philosophy.* VLADIMIR SVOBODA, BJØRN JESPERSEN, COLIN CHEYNE, eds., VARIOUS trans. Dunedin, New Zealand: Otago, 2004. 901 p. Cloth $49.95.

SIMPLICIUS: *On Aristotle's On the Heavens 2.1-9.* IAN MUELLER, trans. Ithaca, NY: Cornell, 2004. ix, 224 p. Cloth $75.00.

VATTIMO, GIANNI: *Nihilism and Emancipation: Ethics, Politics, and Law.* SANTIAGO ZABALA, ed., WILLIAM MCCUAIG, trans. New York: Columbia, 2004. xxv, 197 p. Cloth $24.50.

VIRILIO, PAUL: *Negative Horizon.* MICHAEL DEGENER, trans. New York: Continuum, 2005. viii, 227 p. Cloth $29.95.

WANNENWETSCH, BERND: *Political Worship: Ethics for Christian Citizens.* MARGARET KOHL, trans. New York: Oxford, 2004. xiii, 402 p. Cloth $145.00.

WEBER, ELISABETH, ed.: *Questioning Judaism: Interviews.* RACHEL BOWLBY, trans. Stanford: University Press, 2004. x, 173 p. Cloth $50.00, paper $21.95.

WEININGER, OTTO: *Sex and Character.* DANIEL STEUER and LAURA MARCUS, eds., LADISLAUS LÖB, trans. Bloomington: Indiana UP, 2005. liv, 437 p. Cloth $75.00.

WITTGENSTEIN, LUDWIG: *The Big Typescript: TS 213.* C. GRANT LUCKHARDT and MAXIMILIAN A.E. AUE, eds. and trans. Malden, MA: Blackwell, 2005. xviii, 516 p. Cloth $124.95.

Enlightening

ON BULLSHIT

Harry G. Frankfurt

"[Frankfurt] tries, with the help of Wittgenstein, Pound, St. Augustine and the spy novelist Eric Ambler, among others, to ask some of the preliminary questions—to define the nature of a thing recognized by all but understood by none.... What is bullshit, after all?"
—Peter Edidin, *New York Times*

"A gem of psychological insight, social commentary, philosophical analysis, and good humor.... It is one of the most enjoyable and humanly illuminating short pieces of philosophy produced in the past fifty years."—Raymond Geuss, University of Cambridge

Cloth $9.95 ISBN 0-691-12294-6

PHILOSOPHICAL MYTHS OF THE FALL

Stephen Mulhall

Did post-Enlightenment philosophers reject the idea of original sin and hence the view that life is a quest for redemption from it? In *Philosophical Myths of the Fall*, Stephen Mulhall identifies and evaluates a surprising ethical-religious dimension in the work of three highly influential philosophers: Nietzsche, Heidegger, and Wittgenstein.

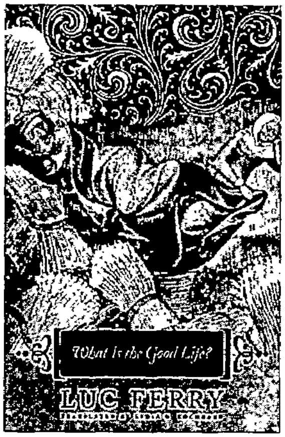
"This book is extremely intelligent, genuinely original, and very well written. Mulhall's suggestion that Nietzsche, Heidegger, and Wittgenstein 'want to preserve a recognizable version of the Christian conception of human nature' is very intriguing indeed, and he develops it splendidly."—Richard Rorty, Stanford University

Princeton Monographs in Philosophy
Harry Frankfurt, series editor
Cloth $29.95 ISBN 0-691-12220-2 Due August

100 *Celebrating 100 Years of Excellence*
PRINCETON
University Press
800-777-4726
Read excerpts online
www.pup.princeton.edu

philosophy

NEW FROM CHICAGO



WHAT IS THE GOOD LIFE?
Luc Ferry
Translated by Lydia G. Cochrane
"For anyone who wants to know why French philosophy after Foucault and Derrida continues to be of global importance, Luc Ferry's *What Is the Good Life?* is a must. . . . An ideal introduction to the thought of one of the leading intellectuals of today's France."
—Edmund Leites, Queens College, City College of New York
CLOTH $45.00

FOUCAULT AND THE IRANIAN REVOLUTION
Gender and the Seductions of Islamism
Janet Afary and Kevin B. Anderson
"This book, which contains the first full translation of Foucault's articles on the Iranian revolution, is essential reading for his political positions in general and his conception of power, as well as for his sexual politics."—Fredric Jameson
PAPER $24.00

FORGIVENESS
Vladimir Jankélévitch
Translated by Andrew Kelley
"*Forgiveness* lets us see why Vladimir Jankélévitch is considered one of the greatest moral philosophers of the twentieth century. His dazzling analysis shows us the singular difficulty and moral force of the art of forgiveness. This book should have an influence as lasting as it is profound."—Arnold I. Davidson, editor of *Foucault and His Interlocutors*
CLOTH $29.00

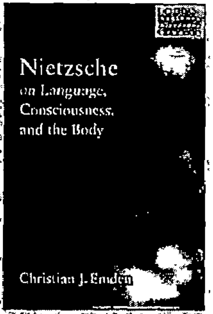
Now in paper

CAMUS AND SARTRE
The Story of a Friendship and the Quarrel that Ended It
Ronald Aronson
"A masterful synthesis of intellectual history, political context, and biographical narrative . . . A book that will reward both those unfamiliar with either thinker and the expert. It will doubtless be the standard account of the Sartre-Camus debate for a long time to come."—Scott McLemee, *Bookforum*
PAPER $19.00

READING ZEN IN THE ROCKS
The Japanese Dry Landscape Garden
Francois Berthier
Translated and with a Philosophical Essay by Graham Parkes
"An absolutely stunning short book on the meanings and philosophy of the Japanese Zen rock garden. . . . Berthier, however, with gentle, almost poetic prose, leads the reader into the mysteries of Ryoanji and the art of making a rock garden."—Scott Gordon, *Daily Yomiuri*
PAPER $14.00

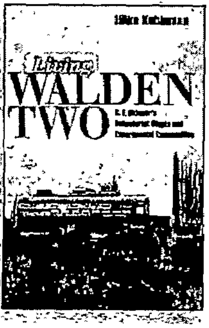
THE UNIVERSITY OF CHICAGO PRESS
1427 East 60th Street, Chicago, IL 60637 • www.press.uchicago.edu

UNIVERSITY OF ILLINOIS PRESS
800-537-5487 · ww.press.uillinois.edu

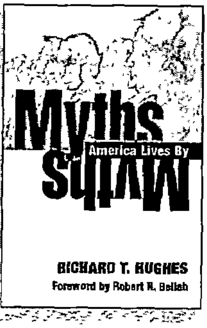


Nietzsche on Language, Consciousness, and the Body
CHRISTIAN J. EMDEN
This work takes a historical approach to Nietzsche's work on language, connecting it to his predecessors and contemporaries rather than his successors.
International Nietzsche Studies
Hardcover, $35.00

Thinking through Kierkegaard
Existential Identity in a Pluralistic World
PETER J. MEHL
A critical evaluation of Søren Kierkegaard's vision of the normatively human, of who we are and might aspire to become, and of what Mehl calls our existential identity.
Hardcover, $35.00



Living Walden Two
B. F. Skinner's Behaviorist Utopia
and Experimental Communities
HILKE KUHLMANN
An entertaining tour of the dynamics of the nearly three dozen communities inspired by *Walden Two,* with an in-depth examination of the two surviving Skinnerian communities: Comunidad Los Horcones in Mexico, and Twin Oaks in Virginia.
Illustrated. Hardcover, $35.00



Myths America Lives By
RICHARD T. HUGHES
Foreword by Robert N. Bellah
Hughes identifies the five key myths that lie at the heart of the American experience, showing that by canonizing these seemingly harmless myths of national identity as absolute truths, America risks undermining the sweepingly egalitarian promise of the Declaration of Independence.
New in Paperback, $19.95

New books from The MIT Press

Sweet Dreams

Philosophical Obstacles to a Science of Consciousness

Daniel C. Dennett

The author of *Consciousness Explained* revises and renews his Multiple Drafts Model of consciousness in the light of recent research.

The Jean Nicod Lectures • A Bradford Book
216 pp., 3 illus. $28

Onflow

Dynamics of Consciousness and Experience

Ralph Pred

"Ralph Pred has written a brilliant book. He synthesizes the work of James and Whitehead with contemporary work in philosophy and neurobiology to produce a powerful and original account of conscious experience." — John Searle, University of California, Berkeley

A Bradford Book • 368 pp. $35

The Old New Logic

Essays on the Philosophy of Fred Sommers

edited by David S. Oderberg
foreword by P. F. Strawson

A diverse group of contributors reflect on the philosophical legacy of Fred Sommers and his efforts to revive and refashion traditional Aristotelian logic for a post-Fregean world.

A Bradford Book • 256 pp., 16 illus. $35 paper

now in paperback

Transcritique

On Kant and Marx

Kojin Karatani
translated by Sabu Kohso

"An immensely ambitious theoretical edifice in which new relations between Kant and Marx are established, as well as a new kind of synthesis between Marxism and anarchism."
— Fredric Jameson, Duke University, author of *Postmodernism, or, The Cultural Logic of Late Capitalism*

336 pp. $18 paper

Adapting Minds

Evolutionary Psychology and the Persistent Quest for Human Nature

David J. Buller

"Buller's critique of evolutionary psychology is measured, logical, and clearly developed. . . .This is philosophy of science in the trenches, and it is excellent." — Elliott Sober, University of Wisconsin, Madison

225 pp. $34.95

Reconstructing the Cognitive World

The Next Step

Michael Wheeler

"This book has the potential to radically reshape how cognitive science and analytical philosophy view the mind."
— Mark A. Bedau, Reed College, Editor-in-Chief of *Artificial Life*

A Bradford Book • 432 pp. $35

Panpsychism in the West

David Skrbina

"Provides a long overdue and much needed reexamination of this age-old doctrine, one which still retains some fascination for modern philosophy of mind." — William Seager, University of Toronto at Scarborough

A Bradford Book • 336 pp. $35

now in paperback

Terms and Truth

Reference Direct and Anaphoric

Alan Berger

"A major work. Berger offers an elegant examination of issues that have been in controversy for the last forty years and that have been and are being discussed by the best philosophers of language."
— Gilbert Harman, Princeton University

A Bradford Book • 232 pp. $20 paper

To order call **800-405-1619.**

http://mitpress.mit.edu

The Rational Imagination

How People Create Alternatives to Reality

Ruth M. J. Byrne

"Brings together an account of rationality and creativity to present a unique and powerful theory of how we think about alternatives to reality."
— Philip N. Johnson-Laird, Princeton University

A Bradford Book • 472 pp., 4 illus. $32

Evolution in Four Dimensions

Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life

Eva Jablonka and Marion J. Lamb
with illustrations by Anna Zeligowski

"With courage and verve, and in a style accessible to general readers, Jablonka and Lamb lay out some of the exciting new pathways of Darwinian evolution that have been uncovered by contemporary research." — Evelyn Fox Keller, MIT, author of *Making Sense of Life: Explaining Biological Development with Models, Metaphors and Machines*

472 pp., 63 illus. $34.95

Economic Theory and Cognitive Science

Microexplanation

Don Ross

"The most important new work in the philosophy of economics in years!"
— Alex Rosenberg, Duke University

A Bradford Book • 384 pp., 5 illus. $45

Ethics of the Body

Postconventional Challenges

edited by Margrit Shildrick and Roxanne Mykitiuk

"A remarkable group of essays that use postmodern thought to deconstruct conventional bioethics, biomedicine, and biotechnology. . . . This book will be a significant contribution to bioethics."
— Rosemarie Tong, University of North Carolina at Charlotte

Basic Bioethics series • 328 pp. $27 paper

Solidarity

From Civic Friendship to a Global Legal Community

Hauke Brunkhorst
translated by Jeffrey Flynn

"'Solidarity' is a much used and abused concept. In his comprehensive overview, Hauke Brunkhorst provides us with the first major treatment of its history and theory. This is a must-read for all interested in moral, political, social, and legal issues." — Seyla Benhabib

Studies in Contemporary German Social Thought series • 336 pp. $42

Zone Books

Echolalias

On the Forgetting of Language

Daniel Heller-Roazen

A lively and far-reaching investigation into the persistence and disappearance of speech, in individuals and in linguistic communities.

Distributed for Zone Books • 288 pp. $28

The Cradle of Humanity

Prehistoric Art and Culture

Georges Bataille
edited and introduced by Stuart Kendall
translated by Stuart and Michelle Kendall

A radically interdisciplinary—anthropological, religious, philosophical, aesthetic—inquiry into the origins of human consciousness, community, and potential.

Distributed for Zone Books
224 pp., 15 illus. $28

now in paperback

Pure Immanence

Essays on A Life

Gilles Deleuze
translated by Anne Boyman
introduction by John Rajchman

"A glimpse into some seminal themes of [Deleuze's] work. This book is worth purchasing for the Nietzsche essay alone." — Jennifer Matey, *Metapsychology*

Distributed for Zone Books
102 pp. $16 paper

CAMBRIDGE

BEST IN SCHOLARSHIP FROM CAMBRIDGE

Human Identity and Bioethics
David DeGrazia
$70.00: Hardback: 0-521-82561-X: c.270pp
$25.99: Paperback: 0-521-53268-X

Values of Beauty
Historical Essays in Aesthetics
Paul Guyer
$70.00: Hardback. 0-521-84490-8: 384pp
$27 99: Paperback: 0-521-60669-1

Metaphor in Culture
Universality and Variation
Zoltán Kövecses
$75.00: Hardback: 0-521-84447-9: 332pp

Now in paperback!

The End of Art
Donald Kuspit
$18.99: Paperback: 0-521-54016-X: 224pp

Ethics Done Right
Practical Reasoning as a Foundation for Moral Theory
Elijah Millgram
$70.00: Hardback: 0-521-83943-2: 360pp
$27.99: Paperback: 0-521-54826-8

Laws of Fear
Beyond the Precautionary Principle
Cass R. Sunstein
The Seeley Lectures 6
$65.00: Hardback: 0-521-84823-7. 246pp
$23.99. Paperback: 0-521-61512-7

Medicine and Philosophy in Classical Antiquity
Doctors and Philosophers on Nature, Soul, Health and Disease
Philip J. van der Eijk
$95.00: Hardback: 0-521-81800-1· 418pp

Hegel's Philosophy of Reality, Freedom, and God
Robert M. Wallace
Modern European Philosophy
$75.00: Hardback: 0-521-84484-3: 384pp

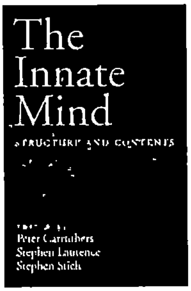
Legal Reason
The Use of Analogy in Legal Argument
Lloyd L. Weinreb
$60.00: Hardback. 0-521-84967-5: 192pp
$18.99: Paperback: 0-521-61490-2

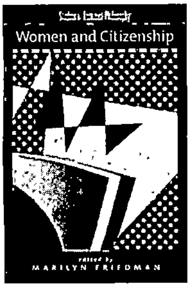
Prices subject to change.

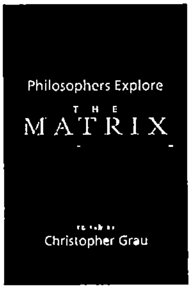
www.cambridge.org/us/philosophy/

CAMBRIDGE
UNIVERSITY PRESS

New from
Oxford







Objects of Metaphor

SAMUEL GUTTENPLAN

This book offers novel conceptions of the relationship between simile and metaphor, the notion of dead metaphor, and the idea of metaphor as a robust theoretic kind. Without denying that metaphor can sometimes be merely ornamental, Guttenplan justifies the view of metaphor as fundamental to language and the study of language.

July 2005 paper $35.00 cloth $90.00

The Innate Mind

Structure and Contents

Edited by PETER CARRUTHERS, STEPHEN LAURENCE, and STEPHEN STICH

"It's not just a survey of the terrain, though it is that; it also contains a number of papers that push the boundaries and make major new contributions to several hot debates in cognitive science...I predict that the book will become a must-have collection."—Fiona Cowie, California Institute of Technology

July 2005 paper $35.00 cloth $74.00

Women and Citizenship

Edited by MARILYN FRIEDMAN

Containing essays by a well-known group of scholars, including Iris Marion Young, Alison Jaggar, Martha Nussbaum, and Sandra Bartky, this book examines the conceptual issues and strategies at play in the feminist quest to give women full citizenship status.

July 2005 paper $24.95 cloth $74.00

Philosophers Explore The Matrix

CHRISTOPHER GRAU

"This book is a rare thing: an anthology on a highly popular topic that actually includes many intellectually excellent contributions, and is at the same time engaging and intelligible to the non-academic public."—Ned Block, New York University

August 2005 paper $24.00 cloth $74.00

Towards Non-Being

The Logic and Metaphysics of Intentionality

GRAHAM PRIEST

Graham Priest presents a ground-breaking account of the semantics of intentional language—verbs such as "believes," "fears," "seeks," or "imagines." *Towards Non-Being* proceeds in terms of objects that may be either existent or non-existent, at worlds that may be either possible or impossible.

July 2005 $49.95

Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

JKC.042705

COLUMBIA UNIVERSITY PRESS

REVISED EDITION

Rules and Representations

Noam Chomsky

Foreword by Norbert Hornstein

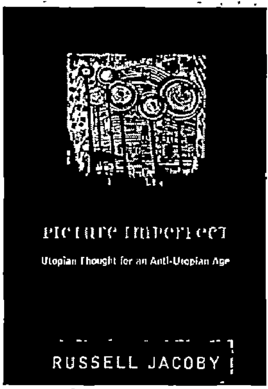
Based on Chomsky's 1978 Woodbridge Lectures, this seminal work in linguistics and the study of the human mind is now available in a revised edition with two new essays.

The Long Arc of Justice

Lesbian and Gay Marriage, Equality, and Rights

Richard D. Mohr

Engaging the whole spectrum of public-policy issues affecting gays and lesbians from a humanistic and philosophical approach, Richard Mohr assesses the logic and ethics of gay rights.



Picture Imperfect

Utopian Thought for an Anti-Utopian Age

Russell Jacoby

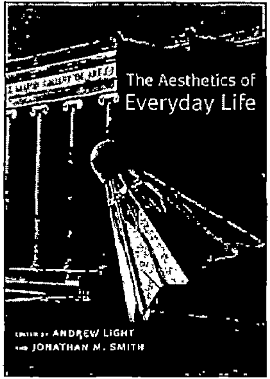
Considering the works of Theodor Adorno, Walter Benjamin, Ernst Bloch, and others, Jacoby argues that a return to an iconoclastic utopian spirit is vital for reviving society's dormant political imagination.

Decameron and the Philosophy of Storytelling

Author as Midwife and Pimp

Richard Kuhns

Kuhns approaches *Decameron* from a variety of literary, philosophical, and historical perspectives, revealing the ways in which its comic and sexual elements lead to philosophical debate and moral argument.



The Aesthetics of Everyday Life

Andrew Light and Jonathan M. Smith, eds.

This imaginative new collection explores the aesthetic qualities of human relationships, sports, food, and environment, offering an important expansion of traditional aesthetics.

columbia.edu/cu/cup

Patta
1.3.06

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NUMBER 8
AUGUST 2005

page

381 *A Plea for Things That Are Not Quite All There: Or, Is There a Problem about Vague Composition and Vague Existence?* Nicholas J.J. Smith

COMMENTS AND CRITICISM

422 *Doors, Keys, and Moral Luck: A Reply to Domsky* Daniel Statman

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Christopher Peacocke, Philip Pettit, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.

You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2005

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.

POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2005 by the Journal of Philosophy, Inc.

ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NO. 8, AUGUST 2005

A PLEA FOR THINGS THAT ARE NOT QUITE ALL THERE: OR, IS THERE A PROBLEM ABOUT VAGUE COMPOSITION AND VAGUE EXISTENCE?

Although they might not express themselves in quite this way, nonphilosophers tend to think that *mereological composition is a vague matter*: sometimes it occurs, sometimes it does not, and sometimes it *sort of* occurs. For example, when I am building a boat, at first the timbers that I have acquired for the job do *not* jointly compose an entity; in the end they do—they compose the boat that I have built; and in between they *sort of* or *more or less* or *to some extent* compose an entity, which in turn *sort of* or *more or less* or *to some extent* exists—this entity being the boat I am building. This idea seems innocuous enough. However, the orthodox view among philosophers is that composition can never be a vague matter, because vague composition entails vague existence (the common-sense view agrees with this step), and vague existence is impossible if not nonsensical. Let us call the following the *orthodox argument*:[1]

O1: Whenever there is a case of vague composition, there is a case of vague existence.
O2: There are no cases of vague existence (because vague existence is at least impossible, if not nonsensical).
OC: Therefore there are no cases of vague composition.

---

* I am grateful to Ed Mares and Huw Price for their encouragement and support of this project. Versions of the paper were presented at the Annual Conference of the Australasian Association of Logic in November 2002, and at the Australian National University in January 2005; I thank the audiences on those occasions for their comments and questions. Finally, thanks to the anonymous referees for their detailed comments.

[1] We may attribute this argument to "Davore Lewder" (the fusion of David Lewis and Theodore Sider). The actual arguments of Lewis, *On the Plurality of Worlds* (New York: Blackwell, 1986), pp. 212–13, and Sider, *Four-Dimensionalism: An Ontology of Persistence and Time* (New York: Oxford, 2001), pp. 126–27, will be discussed in section VI below.

0022-362X/05/0208/381–421

The validity of the argument being beyond dispute, there are three approaches we can take: deny one (or both) of the premises, or accept the conclusion.

Accepting OC. If composition is never vague, we have three options: sharp boundaries, nihilism, and unrestricted composition. The *sharp boundaries* view has it that in a series of cases, each of which is extremely similar to its successor and/or predecessor in all respects relevant to whether or not composition occurs, there are adjacent cases such that in one, composition occurs, while in the other, it does not. For example, start at case 0 with a chair with *n* parts; move to case 1 by moving part 1 one nanometer from the others;[2] move to case 2 by moving part 2 one nanometer from the others; ... move to case $n + 1$ by moving part 1 another nanometer from the others; and so on.... Continue the series until all the parts are widely separated from one another. According to the sharp boundaries view, at some particular point in the series, the chair suddenly ceases to exist. *Nihilism* is the view that composition *never* occurs: there are no composite objects. *Unrestricted composition* is the view that composition *always* occurs: for any bunch of objects whatsoever, there exists an object that is composed of just the objects in the bunch. The sharp boundaries view has won few converts.[3] Nihilism is doing slightly better: once defended by Peter Unger[4] and rejected by almost everyone else, it has recently been reinvigorated by Gideon Rosen and Cian Dorr.[5] Unrestricted composition is doing best of all, having been famously and influentially defended by David Lewis, and more recently defended by (among others) Michael Rea[6] and Theodore Sider.

Three ways, then, to accept OC: and each horribly implausible! In philosophy, familiarity tends to breed acquiescence rather than contempt, but the reader should not need much reminding that each of these views is intuitively repulsive. During the demolition process, my house existed right up until the point when a particular board was removed, at which time it ceased to exist entirely?! My house does not exist at all (only timbers arranged house-wise)?! Not only my house exists, but also another object, composed of my house together with Phar Lap's heart and Australia II's winged keel?! I do not mean to suggest that these incredulous questions constitute any sort of *argument* against the views in question—I simply wish to remind the reader that each of these views is prima facie absurd. So why has acceptance of OC become orthodox? Because so many philosophers have been convinced that they cannot have what pre-philosophical intuition desires: a vague restriction on composition. They have been convinced of this because they accept the orthodox argument. But can we deny one of the argument's premises?

---

[2] For ease of exposition, I assume the parts are numbered.

[3] See Sider, p. 124, for an argument against the view.

[4] "I Do Not Exist," in Michael C. Rea, ed., *Material Constitution: A Reader* (Lanham, MD: Rowman and Littlefield, 1997), pp. 175–90.

[5] "Composition as a Fiction," in Richard M. Gale, ed., *The Blackwell Guide to Metaphysics* (Malden, MA: Blackwell, 2002), pp. 151–74.

[6] "In Defense of Mereological Universalism," *Philosophy and Phenomenological Research,* LVIII (1998): 347–60.

Denying O1. Why think that vague composition entails vague existence? Well, suppose there are three teacups in the cupboard. Do they jointly compose a fourth entity, Cup? Suppose they do not. It is still logically possible that Cup exists, even though the three cups do not compose Cup; but once we abandon the claim that the cups compose Cup, we have no reason at all to believe in Cup any more. Cup seems to exist only to the extent that the cups compose Cup. If the cups *sort of* compose an entity Cup, this is not a matter of (a) there being four entities, with a vague relation holding between one of them and the other three; rather, it is a matter of (b) there being three definite entities, and one shady one which only *sort of* exists. Of course, (b) is not the only consistent way of thinking of matters in this area—(a) is a *consistent* option—but it is by far the most natural way. If we want to defend an intuitively satisfying view of composition, we should not reject O1.[7]

Denying O2. Can we then make sense of the idea of vague existence? The prevailing attitude among philosophers is that we cannot. The aim of this paper, however, is to show that we can: my aim is to *make good sense* of vague composition, vague existence, and the relations between them. I take the state of play to be as follows. What we would have *liked* to be able to say, when we came to philosophy, is that composition is sometimes vague, and that when it is, existence is vague, too. However, as philosophers, we can see no way of making clear sense of this idea, and so we abandon it in favor of one or other way of accepting OC. But in fact, the original idea makes perfectly good sense—or so I shall argue here. My aim is to make this view a serious contender among theories of composition; as for why we should actually accept the view to be presented here, once it has been clearly stated and shown to make good sense, I take the reason to be simply that it is what we wanted all along, but thought we could not have![8] Of course, once the neglected view is on the table and receives serious consideration, it may turn out that there is some subtle reason why it should be rejected in favor of a view that has *less initial* appeal. But such considerations are premature here: for now, what I want to show is that—contrary to received wisdom—the neglected view *makes perfectly good sense.*

---

[7] I am not aware of anyone who argues that vague composition does not entail vague existence. Michael Morreau, "What Vague Objects Are Like," this JOURNAL, XCIX, 7 (July 2002): 333–61, argues that *parthood* is vague; but that *neither* composition nor existence is vague. A referee commented that the link between vague composition and existence is not straightforwardly intuitive, for instead of saying that "insofar as the timbers sort of, or more or less, compose the boat, the boat sort of, or more or less, exists...one might as well say that insofar as the timbers sort of, or more or less, compose a boat, they sort of, or more or less, fall into the extension of the predicate 'boat'." Consider the claim that they (that is, the timbers) fall into the extension of the predicate 'boat'. This might mean one of two things. First, it might mean that each of the individual timbers falls into the extension of 'boat'. However, no one would be tempted to say this, so let us set this reading aside. Second, it might mean that the sum of the timbers falls into the extension of 'boat'. However, we cannot claim this—or deny it—unless we assume that the sum of the timbers exists. But ordinary persons will hesitate over precisely this question of whether the sum of the timbers exists at all, when the timbers are only partially assembled. It thus seems that the question of existence, and the question of predication, are not on a par. Ordinary persons will hesitate over both, but their hesitation over the existence question is first order (it is simply unclear, when the boat is half built, whether the timbers compose anything), whereas their hesitation over the predication question is second order (any answer to the question presupposes an affirmative answer to the existence question—a question over which they are already hesitating).

In what follows, I shall be employing fuzzy logic/set theory as my working framework for handling vagueness. I employ the fuzzy framework here because I think it is the correct framework for the treatment of vagueness.[9] This is a minority view among contemporary philosophers. To those who oppose this view, I respond first that I defend the fuzzy approach to vagueness at length elsewhere, and do not wish to repeat myself here, and second that everything that I do in this paper could also be done within the framework of a linguistic theory of vagueness—in particular, within the *degree-theoretic* form of supervaluationism.[10] I leave the details of such alternative formulations to readers who have sympathy with the positive views of this paper, but not with the background theory of vagueness employed here.

---

[8] Lewis himself admits as much, noting that "To restrict composition in accordance with our intuitions would require a vague restriction": he argues that we cannot have such a restriction; but he does not deny that a vague restriction is what intuition desires. Note that Lewis does not merely say that *if* we are to have a restriction, then intuitively, it should be a vague one: he also notes that intuitively, we want a restriction: "We are happy enough with mereological sums of things that contrast with their surroundings more than they do with one another; and that are adjacent, stick together, and act jointly. We are more reluctant to affirm the existence of mereological sums of things that are disparate and scattered and go their separate ways" (*op. cit.*, p. 212).

[9] More subtly, I think it is the best of the major current theories. In my "Vagueness and Blurry Sets," *Journal of Philosophical Logic*, XXXIII (2004): 165–235, I present a new theory of vagueness which is (in a way explained clearly in that paper) of the same abstract kind as the fuzzy theory, while being significantly different in its concrete details—and it is *this* theory that I think is the correct framework for the treatment of vagueness. The views of the present paper could be recast within this new theory of vagueness—but I do not do this here because the views of the present paper and the new theory of vagueness are not a package deal: as discussed further in the text below, one can develop the views of the present paper within a variety of background frameworks for handling vagueness.

[10] For this view, see Hans Kamp, "Two Theories about Adjectives," in Edward L. Keenan, ed., *Formal Semantics of Natural Language* (New York: Cambridge, 1975), pp.

Before presenting the details of my proposal, I shall present a sketch of the overall picture.

First thought: intuitively, there is no such thing as the mereological fusion of the Empire State Building (E) and the Sydney Opera House (S): there simply is no object which has E and S as parts, and which is such that any of its parts overlaps one of E or S. However, there *could be* such a thing: just transport S and attach it to E (in an architecturally coherent and structurally sound way) and such an object would exist. So the sum of E and S is a possible object, but not an actual one. More generally, for any bunch of objects at all, their mereological fusion is a *possible* object, but in general does not *actually* exist.

Second thought: when I build a boat, I begin by acquiring the timbers that I shall need, and I stack them in numerous piles in my shed, garage, and garden. At this stage, the timbers do not jointly compose a further object. When the job is complete, the timbers do compose a further object: my boat. In between, as the timbers are assembled and fastened together, they gradually come to compose a further thing, and this further thing gradually comes into existence.

Combining these two thoughts: (1) I start with a picture familiar from the semantics of quantified modal logic, in which we have an overall domain D of all possible objects, and a subset D_w of this domain for each possible world w, consisting of the things which exist *at world* w. I alter the picture as follows: D remains a crisp set, but each D_w becomes a *fuzzy* subset of D—this is where the idea of vague existence comes in (it can be a matter of degree whether a certain object exists at a certain world). (2) I distinguish *two* parthood relations: *notional* and *concrete*. Notional mereology is classical extensional mereology with unrestricted composition. In accordance with the first thought above, this operates on the background set D of all possible objects: for any bunch of possible objects, there is a possible object which is the (notional) fusion of that bunch. Concrete parthood, on the other hand, is vague. It has a lot to do with contact and adhesion. For example, a given plank is a *concrete* part of my boat to the extent that it is fastened to the rest of the boat. At first, the plank is not a concrete part of my boat at all, but as I tack it in position, then glue it, then

---

123–55; "The Paradox of the Heap," in Uwe Mönnich, ed., *Aspects of Philosophical Logic* (Boston: Reidel, 1981), pp. 225–77, see pp. 234–35; Lewis, "General Semantics," in his *Philosophical Papers*, Volume I (New York: Oxford, 1983), pp. 189–232, see pp. 228–29, and "Survival and Identity," in his *Philosophical Papers*, Volume I, pp. 55–77, see pp. 69–70. Briefly, the idea is this: start with the (ordinary) supervaluationist's admissible interpretations—thus accepting the fundamental supervaluationist idea that vagueness is a linguistic matter, an extreme form of ambiguity—but then instead of saying that a sentence is (super)true if it is true on all admissible interpretations (and so on), say that a sentence's degree of (super)truth is the size (as given by a normalized measure function) of the set of its admissible interpretations on which it is (classically) true.

screw it, and then the glue cures, it gradually becomes a concrete part of my boat to a greater and greater degree, until finally it is a definite concrete part. (3) A notionally composite object (an object in D with more than one notional part) fades into/out of concrete existence (at a given world) as its notional parts become concrete parts to a greater/lesser degree (at that world). Thus, consider ES, the notional fusion of E and S. It does not actually exist, but as S is moved to New York and placed on top of E, and the process of joining them progresses—that is, as S and E, the *notional* parts of ES, become *concrete* parts of ES to a greater degree—ES comes gradually into (actual, concrete) existence; in other words, it becomes a member of the domain of the actual world (which is a fuzzy subset of D) to a greater degree. Or again, consider the notional fusion B of the planks which will eventually make up my boat. Before work begins, B does not actually exist. As B's notional parts are assembled and fastened together—that is, as they become concrete parts of B to a greater degree—B comes gradually into actual, concrete existence, until in the end it is a definite member of the domain of the actual world.

The paper proceeds as follows. Sections I, II, and III present the core of my view: they discuss, respectively, vague existence, notional mereology, and concrete mereology. Section IV discusses how to extend the core view to handle mereologically variable objects (objects which change their parts over time). Sections V and VI respond to key objections to my view, and section VII briefly discusses some possible extensions and augmentations of the view. Finally, the appendix discusses some further technical details of my treatment of vague existence.

I. VAGUE EXISTENCE

One standard approach to the semantics of quantified modal logic involves an overall domain D of possible objects, and for each possible world w, a domain D_w consisting of the objects which exist in w. Put most basically, the way I propose to handle vague existence is to follow this approach, but make D_w a *fuzzy* subset of D, rather than a regular (classical, crisp) subset. This has the effect that we can say that some possible object (such as the boat I am building) exists to an intermediate degree in the actual world—and to an increasing degree as work progresses, until finally the boat exists definitely (utterly, fully, to degree 1) in the actual world.

I begin by recalling the classical modal model theory which I shall take as my starting point. We consider a standard first order language, with identity, and with standard modal vocabulary, including an existence predicate E.[11] A classical modal model $\mathcal{M}$ is a five-tuple $(W,R,$

---

[11] More precisely, the symbols of the language are: the unary propositional connectives $\neg$ and $\Box$, and the binary propositional connective $\vee$; the quantifier $\forall$; the parentheses (and); infinitely many variables $x_1, x_2, \ldots$; infinitely many predicate letters A^n_k, for

$D,Q,V)$. W is a set of worlds. R is a binary relation on W (the accessibility relation). D is a set of objects (the domain of all possible objects). Q is a function which assigns to each $w \in W$ a subset D_w of D (D_w is the set of things that exist in world w). V is a valuation function which assigns to each n-ary predicate an extension (a set of n-tuples of members of D) in each world w.[12] We stipulate that in the case of the identity predicate, V assigns it the real identity relation on D (at every world), and in the case of the existence predicate, V assigns it D_w at w. A *value assignment* based on a model $\mathcal{M} = (W,R,D,Q,V)$ is a function ν which assigns a member of D to every variable x. $V_v^M(\alpha,w)$ is the truth value of the wf α at the world w in the model $\mathcal{M}$ when the individual variables are given the values assigned to them by ν. This truth value is either 1 (true) or 0 (false). The truth value of any wf at any world is given by the following clauses:[13]

- $V_v^M(Px_1,...,x_n,w) = V^M(P,w)(v(x_1),...,v(x_n))$
  that is, $Px_1...x_n$ is true at w iff $(\nu(x_1),...,v(x_n))$ is in the extension of P at w.
- $V_v^M(\neg\alpha,w) = 1 - V_v^M(\alpha,w)$
  that is, $\neg\alpha$ is true at w iff α is false at w.
- $V_v^M(\alpha\vee\beta,w) = \bigvee\{V_v^M(\alpha,w), V_v^M(\beta,w)\}$
  that is, $\alpha\vee\beta$ is true at w iff α or β is true at w.
- $V_v^M(\Box\alpha, w) = \bigwedge\{V_v^M(\alpha,w'): \forall w' \text{ with } wRw'\}$
  that is, $\Box\alpha$ is true at w iff α is true at all worlds accessible from w.
- $V_v^M(\forall x\alpha,w) = \bigwedge\{V_\rho^M(\alpha,w): \forall\rho \text{ such that } \forall y \neq x,\ \rho(y) = v(y)\}$
  that is, $\forall x\alpha$ is true at w iff α is true at w relative to all assignments which differ from ν at most in what they assign to x.

This semantics countenances a domain for each world, with each of these domains being a subset of an overall domain of possibilia; yet the quantifiers $\exists$ and $\forall$ range over the entire domain of possibilia, rather than being restricted, at each world, to the domain of that world. Although not a new combination of features—a system of this sort is presented, in favorable terms, by George Hughes and Max Cresswell[14]—

$n,k \geq 1$ (superscripts represent number of arguments; subscripts are index numbers); the binary identity predicate =; and the unary existence predicate E. Well-formed formulas are defined as follows: if $x_1,...,x_n$ are variables and P is an n-ary predicate then $Px_1,...,x_n$ is a wf; if $\mathcal{A}$ and $\mathcal{B}$ are wfs and y is a variable, then $(\neg\mathcal{A}),(\mathcal{A}\vee\mathcal{B}),(\Box\mathcal{A})$ and $(\forall y(\mathcal{A}))$ are wfs; nothing else is a wf. In practice, parentheses will be omitted when this will cause no confusion. The following are defined symbols: $\exists y\mathcal{A} := \neg\forall y\neg\mathcal{A}$; $\mathcal{A}\rightarrow\mathcal{B} := \neg\mathcal{A}\vee\mathcal{B}$; $\mathcal{A}\wedge\mathcal{B} := \neg(\neg\mathcal{A}\vee\neg\mathcal{B})$; $\Diamond\mathcal{A} := \neg\Box\neg\mathcal{A}$; $\Pi y\mathcal{A} := \forall y(Ey\rightarrow\mathcal{A})$ (actualist universal quantification); $\Sigma y\mathcal{A} := \exists y(Ey\wedge\mathcal{A})$ (actualist existential quantification).

[12] Formally, we can think of V as mapping each pair (P,w) of an n-ary predicate P and a world w, to the characteristic function of a subset of D^n, that is, to a function from the set of n-tuples of members of D to $\{0,1\}$.

[13] $\bigvee$ and $\bigwedge$ are, respectively, the supremum (least upper bound) and infimum (greatest lower bound) operations on the reals (they map sets of real numbers to real numbers).

[14] *A New Introduction to Modal Logic* (New York: Routledge, 1996), pp. 303–04.

it is not a common one. It is, however, a highly attractive combination. The simplest semantics for quantified modal logic has one overall domain of possibilia, over which the quantifiers $\exists$ and $\forall$ range, and has no separate subdomains for each world. Given some standard further details, this semantics validates the Barcan Formula ($\Diamond\exists x\Phi \rightarrow \exists x\Diamond\Phi$). Those with actualist tendencies find this objectionable, because they think (to borrow an example from Timothy Williamson[15]) that while it is true that Wittgenstein could have had a son ($\Diamond\exists xPx$), it is not true that there exists something which could have been Wittgenstein's son ($\exists x\Diamond Px$). Saul Kripke[16] presented models with a domain for each world, in which the quantifiers $\exists$ and $\forall$ range, at each world, only over the domain of that world. This semantics does not validate the Barcan Formula. However, this semantics does not genuinely satisfy the concerns underlying the actualists' opposition to the Barcan Formula—because the possibilia are still there in the model, and quantification over possibilia is a feature of the *metalanguage* in which the Kripke semantics is presented.[17] The fact that the object language quantifiers do not reach the mere possibilia is like a gag order preventing us talking about some toxic waste, rather than a proper clean-up job which actually gets rid of it—the waste is still there, and it is spoken of in the text of the gag order. All is not lost, however, for closer inspection reveals that the waste is not toxic at all—that is, mere possibilia are not objectionable entities. In particular, they are *not* the Meinongian horrors of which Quine[18] sought to keep our doorways free. As both Linsky and Zalta, and Williamson make clear, Wittgenstein's possible son is not a shadowy entity who *is* a man, a miner, in therapy, and so on, but an entity which *could have been* a man, a miner, in therapy, and so on. Williamson argues that actualists are *wrong* to wish to rid the modal landscape of such entities, and Linsky and Zalta argue that actualists can *accept* such entities; both papers then argue, on this basis, for a return to the simple semantics, with one overall domain of possibilia, and no separate subdomains for each world. I accept the first part of the argument (the taming of the possibilia[19]), but think the second part (jettisoning the subdomains) is overkill. There are indeed technical advantages in having the quantifiers range over the overall domain,[20] but this does not mean we should get rid of the subdomains for each world. It *simply* means that, at each world, ∃ and ∀ should range over the entire domain, not over that world's subdomain. As mentioned, this *is* a feature of the semantics presented above. Now what of the subdomains? Are they not otiose, if they do not figure in the truth definitions for ∃ and ∀? No. On the contrary, they provide the basis for equipping the modal language with certain essential expressive resources, namely, actualist quantification and an existence predicate. Not only is there an *intuitively* clear distinction, among the overall domain of possibilia, between the ones which exist *around here*, which exist *in this world*, which *actually* exist, which are *part of the furniture of the world*—say it how you will—and the rest, but furthermore, Linsky and Zalta and Williamson *themselves* make use of this very distinction in presenting their accounts (Linksy and Zalta talk of "contingently nonconcrete" objects, and Williamson talks of "mere possibilia"). They thus make a distinction which their own semantics does not recognize. Much better, then, to retreat from Kripke not right back to the simplest semantics (without subdomains), but to the sort of semantics presented above. This gives us actualist quantification (Π,Σ) *and* possibilist quantification (∀,∃), and an existence predicate with which to distinguish the things which are part of the furniture of the world from the rest—with the attendant double advantage of technical tractability *and* expressive power.

---

[15] "Bare Possibilia," *Erkenntnis*, XLVIII (1998): 257–73.

[16] "Semantical Considerations on Modal Logic," *Acta Philosophica Fennica*, XVI (1963): 83–94.

[17] Bernard Linksy and Edward N. Zalta, "In Defense of the Simplest Quantified Modal Logic," *Philosophical Perspectives*, VIII (1994): 431–58, and Williamson, also make this point.

[18] "On What There Is," in his *From a Logical Point of View* (New York: Harper, 1953), pp. 1–19.

[19] It is unnecessary here to enter into the vexed, and largely terminological, dispute over whether possibilia are acceptable *to actualists*.

I now turn to the task of fuzzifying the classical model theory presented above. A fuzzy model is exactly the same as a classical model, except for the following two changes. (i) Q assigns to each w in W a *fuzzy* subset D_w of D (rather than a crisp subset). D_w is still to be interpreted as the set of things that exist in w. This means that a thing may now exist at a world *to an intermediate degree.* (ii) V is a valuation function which assigns to each n-ary predicate a *fuzzy set* of n-tuples of members of D (rather than a crisp subset), in each world w.[21] We still stipulate (a) that in the case of the identity predicate, V assigns it the real identity relation on D (at every world)—*and we assume that real identity is always nonvague*—and (b) that in the case of the existence predicate, V assigns it D_w at w—which, on the other hand, is now a *fuzzy* set.

These two changes are the only ones we make. Thus, in particular, W (the set of all worlds), R (the accessibility relation), and D (the domain of all possible objects) remain nonfuzzy.[22] The notion of a value assignment is also unchanged. As before, $V_v^M(\alpha,w)$ is the truth value of the wf α at the world w in the model $\mathcal{M}$ when the individual variables are given the values assigned to them by v. *The clauses giving the truth conditions of the different sorts of wf are written exactly as above*; however, given the changes just made, the truth values of wfs are now members of [0,1], rather than {0,1}. Thus, for example, the degree of truth of $\alpha\wedge\beta$ at w is the minimum of the degrees of truth of α and β at w, and—because necessity functions like universal quantification over accessible worlds, and universal quantification functions like generalized conjunction—the degree of truth of $\Box\alpha$ at w is the infimum of the degrees of truth of α at all worlds accessible from w. I discuss some further details of fuzzy modal logic in the appendix.

---

[20] See, for example, Hughes and Cresswell, chapters 13–16.

[21] Formally, we can think of V as mapping each pair (P,w) of an n-ary predicate P and a world w, to a fuzzy subset of D^n, that is, to a function from the set of n-tuples of members of D to $[0,1]$.

To sum up the idea of the present section: we have an overall set of all possible objects, over which our existential quantifier $\exists$ ranges. Everything in this set exists *simpliciter*, in the sense of being in the range of $\exists$, and outside of this set, there is simply *nothing at all*. So existence, in the sense of $\exists$, is an all-or-nothing matter. However, for each possible world w, there is a *property* which things in the overall domain of possible objects may or may not possess: the property of existing at world w. (For example, you and I both have the property of existing at the actual world, while Santa Claus does not.[23]) Each of these properties is now a *vague* matter: objects need not possess these properties either outright or not at all—they may possess them to any intermediate degree. This means that when it comes to giving an inventory of the contents of the actual world (or any other world), some of the possible things in the overall domain are definitely in the inventory, and some are definitely out, while others are *sort of* in and *sort of* out. In short, the set of things which exist *in our world* has blurred boundaries.

II. NOTIONAL MEREOLOGY

For any bunch at all of possible things, there is a possible thing which is their fusion. Thus we want to allow ourselves the resources of classical mereology with unrestricted composition—and accordingly, in this section, I present a version of classical mereology. Of course we do not want to assume that the domain of the *actual* world is closed under mereological fusion, but we do wish to assume that the domain of all possible objects is so closed: fusing some actual objects may take us to a nonactual, merely possible object; but fusing some possible objects should never lead beyond the possible objects.

---

[22] Fuzzifying *R* is an interesting idea, but is not relevant for present purposes. This idea is mentioned in Graeme Forbes, "Thisness and Vagueness," *Synthese*, LIV (1983): 235–59, p. 257, n. 12 (where it is attributed to Nathan Salmon), and carried out in Nobu-Yuki Suzuki, "Kripke Frame with Graded Accessibility and Fuzzy Possible World Semantics," *Studia Logica*, LIX (1997): 249–49. Forbes there, and in *The Metaphysics of Modality* (New York: Oxford, 1985), introduces a version of counterpart theory in which the counterpart relation is a matter of degree, but—in contrast to the present paper—the existence-at-a-world relation is *not* a matter of degree.

[23] Sorry, kids.

In presenting our system of mereology, we begin with a primitive (undefined) relation of parthood; we then define various other useful notions in terms of this one; and we then impose some fundamental principles governing this relation.[24]

We have a primitive binary relation of *proper parthood*, symbolized by $\sqsubset$, so $x \sqsubset y$ means that x is a proper part of y. $\sqsubset$ is a *non*fuzzy relation (or a fuzzy relation which always holds to degree 1 or degree 0). We say that x is an *improper* part of y just in case x and y are identical (compare the notion of an improper subset). Thus we define the relation of *proper or improper parthood*, symbolized by $\sqsubseteq$, as follows:

$$x \sqsubseteq y := x \sqsubset y \vee x = y.$$

In future, 'part' (used on its own) will mean proper or improper part. Two objects *overlap* just in case they have a part in common.[25] Thus we define the relation $\boxdot$, where $x \boxdot y$ means that x and y overlap, as follows:

$$x \boxdot y := \exists z (z \sqsubseteq x \wedge z \sqsubseteq y).$$

Two objects are *disjoint* just in case they do *not* overlap. Thus we define the relation $\boxtimes$, where $x \boxtimes y$ means that x and y are disjoint, as follows:

$$x \boxtimes y := \neg x \boxdot y.$$

An object is a mereological *atom* or *simple* just in case it has no proper parts.[26] Thus we define the property A, where Ax means that x is a mereological atom or simple, as follows:

---

[24] The cognoscenti will recognize that not all of the fundamental principles discussed below need be included as axioms in a formal theory of mereology, for some can be derived from combinations of others. For some different axiomatizations of classical mereology see Henry S. Leonard and Nelson Goodman, "The Calculus of Individuals and Its Uses," *Journal of Symbolic Logic*, v (1940): 45–55; Alfred Tarski, "Foundations of the Geometry of Solids," in his *Logic, Semantics, Metamathematics* (New York: Oxford, 1956), pp. 24–29; Alfred Breitkopf, "Axiomatisierung einiger Begriffe aus Nelson Goodmans *The Structure of Appearance*," *Erkenntnis*, XII (1978): 229–47; Peter Simons, *Parts: A Study in Ontology* (New York: Oxford, 1987); and Achille Varzi, "Mereology," in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy* (Fall 2004).

[25] Thus any object overlaps itself, as it has an improper part—itself—in common with itself.

[26] Note the restriction to *proper* parts in this definition: every object has at least one *proper or improper* part, namely itself—so if we did not have this restriction, then *by definition* there would be no atomic objects. Given the definition in the text, it may still of course be the case that *as a matter of fact* there are no atomic objects.

$Ax := \neg\exists y(y \sqsubset x)$.

Given a set S of objects, we say that an object x is a *fusion* or *sum* of S just in case:

(1) everything in S is part of x, that is, $\forall y \in S(y \sqsubseteq x)$; and
(2) every part of x overlaps something in S, that is,
$\forall y(y \sqsubseteq x \rightarrow \exists z \in S(y \boxdot z))$.

Now for our fundamental principles of notional mereology. These fall into four groups.

Partial ordering. First, a minimal requirement on any relation that is to deserve the title 'parthood' is that it partially order its field. Thus we stipulate that the proper parthood relation is *irreflexive* (nothing is a proper part of itself):[27]

N1: $\neg x \sqsubset x$

and *transitive* (if x is a proper part of y and y is a proper part of z, then x is a proper part of z):

N2: $x \sqsubset y \wedge y \sqsubset z \rightarrow x \sqsubset z$.

From these it follows that $\sqsubset$ is *asymmetric* (if x is a proper part of y then y is not a proper part of x). Given N1 and N2 and the definitions above, it follows that the parthood relation $\sqsubseteq$ is reflexive (everything is part of itself), transitive (if x is part of y and y is part of z, then x is part of z) and antisymmetric (if x is part of y and y is part of x, then x and y are identical). Thus the parthood relation is also a partial order.[28]

Supplementation. If a part of an object does not exhaust the object—that is, if it is a *proper* part—then it would seem that there must be *other* proper parts (or at least one, at any rate) which make up the difference. Thus we impose a principle which says that any object which has one proper part also has another proper part, disjoint from the first one:

N3: $x \sqsubset y \rightarrow \exists z(z \sqsubset y \wedge z \boxtimes x)$.

Extensionality. A further requirement is that the system of mereology is extensional, that is, that things with exactly the same proper parts are identical:[29]

---

[27] Note that in what follows, I shall often omit initial universal quantifiers for the sake of increased readability.

[28] $\sqsubset$ is a *strict* partial order and $\sqsubseteq$ is a *nonstrict* partial order; compare $<$ and $\leq$ on the real numbers.

[29] Assuming of course that they have proper parts at all: for we do not want to say that there cannot be two distinct simple objects x and y, that is, objects having no proper parts; yet note that in this case x and y have exactly the same set of proper parts, namely the empty set. Note also that the restriction to proper parts is essential:

N4: $\neg Ax \wedge \neg Ay \wedge \forall z(z \sqsubset x \leftrightarrow z \sqsubset y) \rightarrow x = y$.

Existence of Fusions. The final principle guarantees the existence of a unique mereological sum or fusion for every set of objects:[30]

N5: For every set S, there exists exactly one x such that x is a fusion of S.

In the present section, I have introduced a version of standard classical extensional mereology with unrestricted composition. This does not mean, however, that I have ended up accepting OC (the orthodox conclusion that composition can never be vague). I am committed to the view that *notional composition* can never be vague, but in the next section I shall introduce a new notion of *concrete composition*, which *can* be vague.[31]

III. CONCRETE MEREOLOGY

In the classical mereology of the previous section, for any set of things, there exists a sum of those things—no matter how scattered or disparate those things may be. Ordinarily, however, we would say that my hair is part of me until I cut it off, after which it is part of me no longer: and it is not that we think that the sum of the rest of me and my detached hair exists, but is not me; we think that there simply is no such thing as the sum of me and my detached hair. Considerations of this sort have led many theorists to abandon classical mereology in favor of weaker theories of mereology—in particular, theories in which composition is restricted. I think that a better way to accommodate these considerations is to posit *another* relation as well as $\sqsubset$. Let us refer to $\sqsubset$ as the relation of *notional* proper parthood. We shall now introduce a new primitive binary relation $\lhd$ of *concrete* proper parthood. When I cut off my hair, my hair is no longer a *concrete* part of me: my hair, together with my other parts, still notionally compose a possible entity, but they no longer concretely compose an actual entity. Notional parthood and composition will remain as in the previous section, and the considerations which have led some mereologists to place restrictions on composition will instead find outlet in the treatment of *concrete* parthood and composition. Two points about concrete parthood deserve special mention. First, concrete parthood

---

for any two distinct objects automatically do not have exactly the same set of proper-or-improper parts (for each is a part of itself, but given antisymmetry, both cannot be parts of the other).

[30] Note that I say "every set"—including the empty set—hence giving rise to the fusion of the null set (or "null fusion"). This is a personal preference *and is inessential to what follows.* Note also that if we wanted a first order theory of mereology, we could replace N5 with an axiom schema using a (schematic) unary predicate in place of 'S'.

[31] For further discussion of the objection that I am simply a universalist about composition, see the end of section III.

is a *vague* matter. As one of my hairs slowly falls out, it goes *gradually*, by degrees, from being a part of me to not being a part of me. Second, concrete parthood is not absolute: it varies from world to world (some planks might concretely compose a ship in world *w*, but not in world *w*′, where they remain disassembled); and even within a world, it varies from time to time (and thus the domain of a world varies from time to time, as composite objects come into and go out of existence in that world—but more on this below).

Turning to the second point first: we already have the semantic machinery (section I) to handle relativization of concrete parthood to worlds, but we shall need some new machinery to handle the relativization to times. It is a straightforward matter to augment our models so that sentences are evaluated at world-time pairs rather than simply at worlds. I shall now briefly describe how this may be done; readers uninterested in the details may safely skip the following paragraph.[32]

One thing we do *not* want to assume is that time is exactly the same in all possible worlds: time might begin and end at different points in different worlds; in some worlds time might not end or begin at all; in some worlds it might be discrete, in others dense; and so on. So we shall say that each world w comes equipped with a *time* $\mathcal{T}_w = (T_w, <_w)$, which consists in two things: a set T_w (of time instants), and a binary relation $<_w$ on that set (the earlier-later relation).[33] Now one possibility we would not like to rule out is branching time, that is, the case where one instant has several alternative futures.[34] However, given that we already have possible worlds, we do not need branching time *within* a possible world: each alternative future of a particular time can be confined to its own possible world. Thus we can say that $<_w$ is always *linear* (for any two distinct points of time, one precedes the other), as well as *irreflexive* (no point of time precedes itself) and *transitive* (if point x precedes point y and point y precedes point z, then point x precedes point z). Our semantic definitions of section I undergo the obvious minor changes: (i) The function Q now assigns subsets of D not to each world, but to each world-time pair (w,t) where

---

[32] The reader may wonder why I did not build in time right from the start, in section I. The reason I adopt a two-stage approach is that the framework of fuzzy quantified modal logic introduced in section I is both easier to comprehend initially, and of more *general* interest than the augmented framework which includes times as well as worlds—and with that framework in place, the addition of times is then a simple matter.

[33] I allow the case of the null time: the null set of times together with the null relation thereon.

[34] One possibility we shall *not* countenance, however, is that there might be more than one time dimension in some worlds. On this idea, see, for example, Jack W. Meiland, "A Two-Dimensional Passage Model of Time for Time Travel," *Philosophical Studies*, XXVI (1974): 153–73. (The difference between branching time and two-dimensional time is like the difference between a branching *path* and a *plane*.)

w is a world in W and t is a time in T_w, with the subset $D_{(w,t)}$ assigned being the set of things that exist at time t in world w. (ii) The valuation function V assigns to each n-ary predicate a fuzzy set of n-tuples of members of D, in each world *w at each time t* (in T_w); so V is now a function not from pairs (P,w), but from triples (P,w,t). The clauses giving the truth values for compound wfs are now as follows:[35]

- $V_v^M(Px_1,\ldots,x_n,w,t) = V^M(P,w,t)(v(x_1),\ldots,v(x_n))$
- $V_v^M(\neg\alpha,w,t) = 1 - V_v^M(\alpha,w,t)$
- $V_v^M(\alpha \vee \beta,w,t) = \bigvee\{V_v^M(\alpha,w,t),\ V_v^M(\beta,w,t)\}$
- $V_v^M(\Box\alpha,w,t) = \bigwedge\{V_v^M(\alpha,w',t')\colon \forall w'$ with wRw' and $\forall t' \in T_{w'}\}$
- $V_v^M(\forall x\alpha,w,t) = \bigwedge\{V_\rho^M(\alpha,w,t)\colon \forall\rho$ such that $\forall y \neq x,\ \rho(y) = v(y)\}$

With our extended semantic apparatus in place, we now need some principles governing the concrete parthood relation $\lhd$: principles which are similar enough to our principles governing the notional parthood relation $\sqsubset$ to make $\lhd$ recognizable as a relation of *parthood*, but different enough to accommodate the fact that concrete parthood is a matter of degree and varies from world to world and from time to time, and the fact that concrete composition is not unrestricted. I shall consider our five principles of notional mereology in turn, considering in each case whether we want an analog of the principle for concrete mereology, and if so, what the analog should be. First, however, we impose a principle linking notional and concrete parthood: the principle that nothing can be a positive-degree concrete part of an object, at any world or time, unless it is a *notional* part of that object. Before stating this formally, I need to explain my notation. I employ both an infix notation and a prefix notation, as the context requires. These have distinct, though related, meanings. Infix notation: '$x \lhd y$' is a *sentence*, which has a particular truth value at each world-time (relative to a model and value assignment). Prefix notation: $\lhd_{(w,t)}(x,y)$ is the degree to which x is a concrete part of y at time t in world w; so '$\lhd_{(w,t)}(x,y)$' (which will sometimes be abbreviated to '$\lhd(x,y)$') is a *name* of a real number between 0 and 1 inclusive. (Compare: 'Bob is bald' is a sentence which, let us say, has degree of truth 0.5. 'Bob's degree of baldness' is a name of the number 0.5.) Our linking principle can now be expressed as follows:[36]

---

[35] Note that necessity is interpreted as truth *at all times* in all accessible worlds. A notion of *historical necessity* could easily be introduced, as well as the usual temporal operators *F*, *P*, and so on, but as none of these will play a role in what follows, I shall not discuss them here. For further discussion, see Richmond H. Thomason, "Combinations of Tense and Modality," in D. Gabbay and F. Guenthner, eds., *Handbook of Philosophical Logic*, Volume II (Boston: Reidel, 1984), pp. 135–65.

[36] As in the previous section, initial universal quantifiers will not be written explicitly, but should be assumed.

$$\text{NC: } \neg(x \sqsubset y) \rightarrow \forall w,t[\lhd_{(w,t)}(x,y) = 0].$$

Thus notional parthood is one of the determinants of concrete parthood (but not the only one): nonparts in the notional sense are not even in the running to be parts in the concrete sense.[37]

Partial ordering. Some version of the partial order requirement should govern concrete parthood, generalized to cover the fact that concrete parthood is a vague and nonabsolute (that is, world- and time-relative) relation:

$$\text{C1: Irreflexivity: } \lhd_{(w,t)}(x,x) = 0.$$

This says that the degree to which any object is a concrete part of itself is 0, at every world and every time.

$$\text{C2: Transitivity: } \lhd_{(w,t)}(x,z) \geq \bigwedge\{\lhd_{(w,t)}(x,y), \lhd_{(w,t)}(y,z)\}.$$

This says that at any world and time, the degree to which x is a concrete part of z is greater than or equal to the minimum of the degrees to which x is a concrete part of y and y is a concrete part of z. We do not make the inequality here an equality, because of the following sort of example. Suppose we are building a boat c, and all we have left to do is step the mast. The mast-step a is fastened in place, but the mast b has yet to be fitted to the mast-step. The degree to which a is a concrete part of the mast-step/mast unit ab is 0. The degree to which the mast-step/mast unit ab is a concrete part of the boat c is low, maybe 0 (because only the mast-step part of the mast-step/mast unit is fastened in place). Yet the degree to which the mast-step a is a concrete part of the boat c is 1.

Together with the fact that $0 \leq \lhd(x,y) \leq 1$, C2 implies:[38]

$$\lhd(x,y) = 1 \wedge \lhd(y,z) = 1 \rightarrow \lhd(x,z) = 1.$$

It follows from C1 and C2 that $\lhd$ is asymmetric (that is, $\lhd(x,y) > 0 \rightarrow \lhd(y,x) = 0$).[39] If we define the relation $\unlhd$, where $x \unlhd y$ means that x is a proper or improper concrete part of y, as follows:

$$x \unlhd y := x \lhd y \vee x = y$$

then it follows from C1 and C2 that $\unlhd$ is reflexive (that is, $\unlhd(x,x) =$

---

[37] Thus, in particular, notional simples never have any proper concrete parts at any world-time.

[38] Here and elsewhere, for the sake of increased readability, I omit not only initial universal quantifiers, but also w and t subscripts.

[39] Proof: Suppose that $\lhd(x,y) = m$ and $\lhd(y,x) = n$, with $0 < m,n < 1$; then by transitivity $\lhd(x,x) \geq \bigwedge\{m,n\} > 0$, contradicting irreflexivity.

1),[40] antisymmetric (that is, $\trianglelefteq(x,y) > 0 \wedge \trianglelefteq(y,x) > 0 \rightarrow x = y$)[41] and transitive (that is, $\trianglelefteq(x,z) \geq \bigwedge\{\trianglelefteq(x,y),\trianglelefteq(y,z)\}$).[42]

Supplementation. Do we want a principle which ensures that no object can have one proper concrete part without also having another proper concrete part which is disjoint from the first? There are actually several possible principles here. If we define the relation $\odot$, where $x \odot y$ means that x and y concretely overlap, as follows:

$$x \odot y := \exists z(z \trianglelefteq x \wedge z \trianglelefteq y)$$

then we can consider as possible principles each of the eight claims derived from the following formula by choosing either the condition above the line, or the condition below the line, at each of the three points at which such a choice can be made:

$$\lhd(x,y)\frac{>0}{=1} \rightarrow \exists z\left(\lhd(z,y)\frac{>0}{=1} \wedge \odot(z,x)\frac{=0}{<1}\right)$$

Why might one think that one of these claims is true? Well, one might think that something becomes a concrete part of an object when it is *fastened* to other parts of the object, and that *one* object is not enough for fastening to occur (just as one hand cannot clap, and one person cannot tango)—you need at least two objects, so that they can be fastened *together.* I am not convinced by this, however. For example, suppose that we are building a sculpture on a plinth. The plinth is not part of the sculpture; it is just the base on which the sculpture will sit. Suppose the bottom-most part b of the sculpture is attached to the plinth, but is not attached to any other parts of the sculpture, and those other parts are not attached to each other or to the plinth. In this case, it seems to me that b is a degree-1 proper concrete part of the sculpture, and that the sculpture has no positive-degree concrete parts which are disjoint from b. Thus, I think we do not want an analog of supplementation for concrete parthood.[43]

---

[40] Proof: $x \trianglelefteq y := x \lhd y \vee x = y$, so the degree to which $x \trianglelefteq x$ holds is the maximum of the degree to which $x \lhd x$ holds (which is 0) and the degree to which $x = x$ holds (which is 1), which is 1.

[41] Proof: Suppose $\trianglelefteq(x,y) > 0$ and $\trianglelefteq(y,x) > 0$ and $x \neq y$; then $\lhd(x,y) > 0$ and $\lhd(y,x) > 0$, contradicting asymmetry of $\lhd$.

[42] Proof: Case (i): $x = y = z$; reduces to $1 \geq \bigwedge\{1,1\}$ which is true. Case (ii): $x \neq y \neq z$; reduces to $\lhd(x,z) \geq \bigwedge\{\lhd(x,y),\lhd(y,z)\}$, that is, transitivity of $\lhd$. Case (iii): $x \neq y = z$; reduces to $\lhd(x,z) \geq \lhd(x,y)$, but $y = z$ so $\lhd(x,z) = \lhd(x,y)$, so this is true. Case (iv): $x = y \neq z$; reduces to $\lhd(x,z) \geq \lhd(y,z)$, but $x = y$ so $\lhd(x,z) = \lhd(y,z)$, so this is true.

[43] A referee expressed the following intuition about the statue example: "It seems to me that a natural thing to say about that example is that after the feet (say) have been attached to the plinth, the sculpture exists *simpliciter* and the feet constitute the whole of the sculpture as it is at that time." This intuition *can* be accommodated within the framework of this paper, via the idea (to be introduced below) that at the time in question, the set of statue parts containing only the feet has degree of importance 1.

Extensionality. Note for a start that we do not want to say that if there is a time in a world such that for all z, the degree to which z is a concrete part of x at that world-time is the same as the degree to which z is a concrete part of y at that world-time, then $x = y$. For example, suppose I am building a dinghy, and I have all the parts of my boat cut and shaped, and numbered in accordance with my plan, so I know where each part is supposed to go. As it happens, I have accidentally cut *two* foredecks, and given them the same number. Let one of these foredecks be a and the other b. I am at the stage of building the boat where all parts have been fastened except for the foredeck. Now, consider the notional fusion A of the set containing all the other boat parts and a, and the notional fusion B of the set containing all the other boat parts and b. For every thing z, the degree to which z is a concrete part of A is the same as the degree to which z is a concrete part of B (note that a is a degree-0 concrete part of both, and so is b), yet we do not want to say that $A = B$.

But do we want to say that if, for *every* time t in every world w and for every z, the degree to which z is a concrete part of x at (w,t) is the same as the degree to which z is a concrete part of y at (w,t), then $x = y$ (provided x and y are not atomic)? That is, do we want to impose the following as a principle of concrete mereology?:

$$\neg Ax \wedge \neg Ay \wedge \forall z,w,t(\lhd_{(w,t)}(z,x) = \lhd_{(w,t)}(z,y)) \rightarrow x = y$$

For this to come out false, we would need two distinct nonatomic objects x and y which had exactly the same concrete parts to exactly the same degrees at every world-time. Now, given the extensionality of notional mereology, if x and y are distinct, one must have a notional part the other lacks; and then given NC and the supposition that x and y have exactly the same concrete parts to exactly the same degrees at every world-time, one of x and y must have a *notional* part that is a *degree-0 concrete* part of it at *every* world-time. Thus, if the principle under consideration is sometimes false, then there is a notional fusion which is never fully realized, at any time in any world: some of its possible parts are never (at any world or time) brought together as concrete parts. Now, this conflicts with the idea of a *possible* thing as something which is *actual at some world*. But this is not a knock-down objection, and we might wish to accept that the claim under consideration can be false—in which case we should not, of course, adopt it as a principle of concrete mereology. On the other hand, we might think this principle can never be false. But even so, we *still* should not adopt it as a principle of concrete mereology. This is because it has nothing to do with concrete mereology as such: it has to do with making sure that our space of possible worlds is big enough and varied enough—and it is no skin off the nose of concrete mereology *per se* if there is in fact just one possible world! So either way, we should not adopt this extensionality claim as a principle of concrete mereology.

Existence of Fusions. Our final principle of notional mereology told us under what conditions there exists a fusion of a given set of objects (the answer was: under *all* conditions—notional composition is unrestricted). What we need now is an analogous principle for concrete mereology. This principle should relate concrete parthood to *actual* or *concrete* existence, that is, *existence at a world-time,* as opposed to existence simpliciter, that is, membership of the overall domain D. For we already know that for any bunch of things, there is a possible object which is their notional fusion: what we now want to know is under what conditions (specified in terms of facts about concrete parthood) this notional fusion *concretely* or *actually* exists at a time in a world.

For an *atomic* object x in D we stipulate that for every world w and every time t, $D_{(w,t)}(x)$ is either 1 or 0 (where $D_{(w,t)}(x)$ is the degree to which x exists at (w,t), that is, the degree to which x is a member of the fuzzy set $D_{(w,t)}$, this set being the domain of the world-time (w,t), that is, the set of objects which actually or concretely exist at (w,t), this sort of existence being potentially a vague matter, as discussed in section I). That is, there cannot be vagueness about whether a simple object concretely exists at a time in a world: it exists fully (definitely, to degree 1), or not at all (to degree 0). As for whether or not a given simple *does in fact* exist at a particular world-time, this is a matter for modal geographers: it does not fall within the scope of the theory of concrete mereology, and so we do not introduce any principles concerning the concrete existence of notional simples. (Compare the fact that in notional mereology, we say nothing concerning what, if any, simples exist: we say only, for example, that for any set of simples, there exists a fusion of that set.)

Nonatomic objects (that is, objects with notional proper parts), on the other hand, fade into/out of concrete existence as their notional parts become concrete parts to a greater/lesser degree. But what exactly is the relationship between concrete parthood and concrete existence? As a first try, we might think that if $y_1,\ldots,y_n$ are the notional proper parts of x, then to say that x concretely exists is to say that y_1 is a concrete part of x, *and* y_2 is a concrete part of x, *and* $\ldots y_n$ is a concrete part of x. This leads to the following principle:

$$(1)\ D_{(w,t)}(x) = \bigwedge_{\{y_i\,:\,y_i \sqsubset x\}} \lhd_{(w,t)}(y_i,x)$$

This says that the degree to which an object exists at a world-time is the infimum of the degrees to which each of its notional parts is a concrete part of it at that world-time. However, it is a consequence of this principle that if one notional part y_i of x is completely destroyed

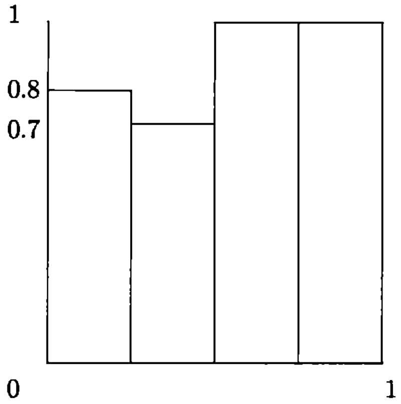


Figure 1

or disconnected at t in w, so that $\lhd_{(w,t)}(y_i,x) = 0$, then x exists to degree 0 at t in w. This is too extreme: a better idea is that where x has n notional simple parts, each simple part can contribute one nth of the concrete existence of x. How much it contributes depends on its degree of concrete parthood of x: if all x's notional parts are degree-1 concrete parts, x concretely exists to degree 1; if one atom is destroyed and becomes a degree-0 concrete part, x loses $^1/_n$ of its concrete existence; if the atom gets somewhat disconnected and becomes a degree-0.5 concrete part of x, x loses $^{0.5}/_n$ of its existence; and so on. This leads us to the following principle:

$$(2)\ D_{(w,t)}(x) = \sum_{i=1}^{n} \lhd_{(w,t)}(y_i, x) \cdot \frac{1}{n}$$

where $y_1,\ldots,y_n$ are all the simple notional proper parts of x (that is, all the y_i such that Ay_i and $y_i \sqsubset x$).

We can add a further subtlety. It might be (for example) that although your transistor radio has nine hundred and eighty parts, some of them are more important than others: some are just for decoration, while others are crucial parts. So it might be that removing one part affects the (actual, concrete) existence of the radio more than removing another part. So we assign each simple notional part y_i of the radio x a *degree of importance to x*, in symbols $\langle y_i\rangle_x$. We allow this degree of importance to vary from world to world and from time to time, that is, we read $\langle y_i\rangle_x$ as $\langle y_i\rangle_{x,(w,t)}$ (although the subscript (w,t) will generally not be written explicitly, for the sake of increased readability). For each y_i and each x, $0 \leq \langle y_i\rangle_x \leq 1$ (at every world-time), and where there are n notional parts $y_1 \ldots y_n$ of x in total, $\Sigma_{i=1}^{n}\langle y_i\rangle_x = 1$ (at each world-time). Thus, we arrive at the following principle:

$$(3)\ D_{(w,t)}(x) = \sum_{i=1}^{n} \lhd_{(w,t)}(y_i, x) \cdot \langle y_i \rangle_{x,(w,t)}$$

where $y_1 \ldots y_n$ are all the simple notional proper parts of x. Note that (2) is simply the special case of (3) where every part has the same degree of importance, that is, $1/n$.

To get a feel for what is going on here, consider figure 1, which depicts a particular notional fusion x at a particular world-time. Each bar of the graph represents one of the notional parts of x. The width of the bar represents the importance of the part to x (at the given world-time), and the height of the bar represents the degree to which that part is a *concrete* part of x (at the given world-time). x's degree of existence (at the given world-time) is represented by the area under the graph: in particular, when all the bars reach to 1, x concretely exists to degree 1. In the case shown in the diagram, we have an object with four notional atomic parts; each part has the same degree of importance; two parts are degree-1 concrete parts, one is a degree-0.8 concrete part, and one is a degree-0.7 concrete part; the degree of concrete existence of the object is $0.25 \times (0.8 + 0.7 + 1 + 1) = 0.875$.[44]

So far our proposal covers only notional fusions that (i) have simple parts (they are not made of gunk), and (ii) have *finitely* many simple parts.[45] Dealing with the second point first, we may generalize the finite, bar-graph case to the infinite case in a standard way. Intuitively, as the number of parts increases to infinity, the bars on the graph become as thin as points, the stepped line across their tops becomes a curve, and we integrate rather than forming a sum.[46] Formally, the details are as follows (they may be skipped by readers so inclined, without jeopardizing their understanding of the rest of the paper).

Consider a notional fusion x. Let S_x be the set of all notional atomic parts of x. Let F_x be a σ-field of subsets of S_x. Let μ be a measure on F_x, and furthermore let μ be normalized, that is, $\mu(S_x) = 1$. μ is to be interpreted as a measure of importance; as before, we allow μ to vary from world to world and from time to time, that is, we read μ as $\mu_{(w,t)}$. Let $f^x_{(w,t)} : S_x \rightarrow [0,1]$ be the function which assigns to each $y \in S_x$ its degree of concrete parthood of x at t in w, that is, $f^x_{(w,t)}(y) = \lhd_{(w,t)}(y,x)$. We require that $f^x_{(w,t)}$ be a measurable function, that is, measurable relative to F_x.[47] Now where $\int$ is the Lebesgue integral, we say:[48]

---

[44] Any reader who finds it absurd to be so precise about the degree of existence of an object is objecting not to the views presented in this paper as such, but to the background framework for the treatment of vagueness: the fuzzy framework. See the discussion in the introductory section above.

[45] There are at least two ways in which the finiteness assumption might fail: an object with finite (or infinite) extension might have infinitely many parts which do not have positive extension; and an object with infinite extension might have infinitely many parts which do have positive extension.

[46] In the finite case the vertical axis represents (roughly speaking) a mass-like quantity, whereas in the infinite case it represents a density-like quantity.

$$(4)\ D_{\langle w,t\rangle}(x) = \int_{S_x} f^x_{\langle w,t\rangle} d\mu$$

---

[47] This constrains our choice of F_x and $f^x_{\langle w,t\rangle}$, but there is no reason to think that this constraint will conflict with natural choices of F_x and $f^x_{\langle w,t\rangle}$. Typically, although perhaps not always, F_x will simply be the set of all subsets of S_x, and then $f^x_{\langle w,t\rangle}$ will automatically be measurable relative to F_x.

[48] A referee thought that the integral should be taken not over the set S_x but over x itself, and that μ should take notional parts of x (rather than sets of such parts) as arguments. The thought could be that there is a *technical* error in my proposal, or that it faces a *conceptual* problem. First, there is no technical error in my proposal. Distinguish the function $f^x_{\langle w,t\rangle}$—which assigns to each notional part of x its degree of concrete parthood of x at t in w—and the measure function μ. We integrate the function f, relative to the measure μ. f should (and does) take parts of x as arguments, and therefore μ should (and does) take *sets* of parts of x as arguments, so that we can talk about the sizes of the preimages under f of given subsets of $[0,1]$ ($[0,1]$ being the codomain of f, and the preimage under f of a given subset of $[0,1]$ being the *set* of parts which get mapped to a value in that subset). (Compare the way in which in probability theory, we have a sample space, a random variable—a function which assigns real number values to *elements* of the sample space—and a probability measure—a function which assigns sizes to *sets* of elements. In determining the expectation of the random variable, we integrate it with respect to the measure.) Second, conceptual issues. Intuitively, μ is supposed to represent the measure of importance of *parts* of x—and yet it assigns values to sets of parts, not parts themselves. This is no deep problem: we regard the degree of importance of an individual *part* as the value which μ assigns to its *singleton*. Shifts of this sort happen all the time when we model intuitive phenomena in a precise way, and while they may cause confusion if we do not note them carefully, once noted they are no cause for concern. On the other hand, there *is* a deep conceptual problem with the alternative proposal that μ take parts, not sets of parts, as arguments. Measures must be additive. On my proposal (where μ takes sets of parts as arguments), μ is additive over unions: the measure of the union of two sets of parts is the sum of the measures of those two sets of parts. On the alternative proposal (where μ takes parts as arguments), the referee suggests—and this would seem to be the only option—that μ be additive over mereological sums: the measure of the mereological sum of a and b is the sum of the measures of a and b. But this yields results which are unacceptable in the context of the present project. My aim is to capture ordinary intuitions about vague composition and existence. One such intuition is that sums of *scattered* concretely-existing objects do not concretely exist (cf. *datum 2* below). Another is that an object which does not concretely exist (that is, its degree of concrete existence is zero) cannot be a concrete part of any other object (that is, to any positive degree—cf. *datum 3* below). The fundamental thought in this paper is that an object concretely exists to the extent to which its notional parts are *concrete* parts of it. Now these ideas will not mesh with the idea that the mereological sum of any two important parts is important—that is, with the idea that μ be additive over mereological sums. For consider, say, a car. The steering wheel and the engine are important parts of the car: the car will not fully concretely exist until they are concrete parts of it. Their sum is a scattered object which does not concretely exist at all: to make it exist we would have to attach the steering wheel directly to the engine. Now, if—as the proposal under consideration implies—their sum is an important part of the car—because it is the sum of two important parts—then the car will not fully exist until this sum is a concrete part of the car; but this cannot happen until the sum concretely exists—and this cannot happen until the steering wheel is directly attached to the engine. So the car cannot fully exist unless the steering wheel is removed from its usual mounting point and attached directly to the engine. This is absurd. Moreover, we clearly face a situation in which the car can *never* fully concretely exist: for whatever sums of its important parts we make concretely exist, we will invariably take others out of concrete existence in the process.

Let us turn now to the first point: the assumption of atomism. In order to accommodate the case of gunk, we have to abandon the restriction to atomic parts. What we do is proceed exactly as above, except that we let S_x be the set of *all* notional parts of x (rather than the set of all notional *atomic* parts of x). This change gives rise to a subtlety; let us consider a specific example. We are building a boat, and it is made of gunk: the planks are not simple notional parts, and nor are the cellulose molecules of which each plank is composed (and so on)—rather, the notional parts go down forever. Now suppose the boat is complete: all planks are fastened in place. We certainly want to say that the boat concretely exists to degree 1 (call this *datum 1*). Consider the notional part *ab* of the boat which consists of two widely separated planks *a* and *b*, say the port shear strake and the starboard garboard strake.[49] This object *ab* has a low—indeed zero—degree of concrete existence (call this *datum 2*), because the notional parts of *ab* have not been brought into contact and fastened together, that is, they are not *concrete* parts of *ab*: we do not want to abandon this claim, for one of our main motivations was to accommodate the intuition that such scattered objects do not actually exist. Now, a very plausible principle is the following: if the degree of concrete existence of an object x at (w,t) is 0, then for any object y, the degree to which x is a concrete part of y at (w,t) is 0 (call this *datum 3*). Now, we face an apparent problem. We are integrating our concrete-parthood function over the set of *all* parts (not just simple ones): *including ab.* But by datum 2, *ab* does not concretely exist, and so by datum 3, *ab* cannot be a concrete part of the boat. Hence, it seems that the integral will not equal 1: some of the notional parts we are integrating over are not concrete parts, so the boat will not fully concretely exist—contradicting datum 1.

In fact, there is no problem here: the three data are actually compatible. We simply say—very plausibly—that the measure of importance at (w,t) of any notional fusion of parts which are *scattered* at (w,t) is zero. Note here that we must be careful to distinguish $\{a,b\}$ and $\{ab\}$ (the former set contains two elements, a and b; the latter set contains one element *ab*, which is the notional fusion of a and b). As both the sets $\{a\}$ and $\{b\}$ have a positive measure of importance, $\{a,b\}$ must also have a positive measure of importance, as measures are additive. But this does *not* mean that $\{ab\}$ must have a positive measure of importance. We are quite free to say that $\{a\}$, $\{b\}$ and $\{a,b\}$ all have positive measure, while $\{ab\}$ has zero measure—and indeed this is eminently plausible. (Imagine a marine engineer inspecting a ship and giving it a positive survey report, and then waking up in a cold sweat in the night thinking 'I checked that the rudder was secure, and also the keel, but I forgot to check their fusion!' The rudder is a crucially important part, so too the keel—but their fusion, *considered as an object in its own right*, is not important at all.) Now, one of the key properties of the Lebesgue integral is that it "ignores" sets of measure zero: two functions which differ in value only over a set of measure zero have the same integral. This means that although *ab* is a concrete part to degree 0, and although *ab* is in the set of parts over which we integrate, this does not affect the degree of concrete existence of the boat: the degree to which *ab* is a concrete part of the boat is entirely irrelevant to the value of the integral—and the same goes for any other notional part of the boat which is a fusion of scattered parts (that is, parts which are scattered at time t in world w).[50]

---

[49] The port shear strake is the topmost plank, adjacent to the deck, on the port side; the starboard garboard strake is the bottommost plank, adjacent to the keel, on the starboard side; so if you imagine the port shear strake and the starboard garboard strake remaining in position (perhaps held by wires) while all other parts of the boat are removed, these planks would be separated by a significant expanse of empty space.

---

[50] A referee raised the following example. "Suppose that object p is a rectangular prism that is the concrete fusion of four nonoverlapping duplicate cubes of homogeneous, atomless, gunk (a, b, c, and d) arranged in a line. Suppose that the domain of the importance measure μ is the set of all notional parts of p. What values should μ assign to a, b, c, and d? It might seem natural to say $\mu(a) = \mu(b) = \mu(c) = \mu(d) = \frac{1}{4}$. But then what value is μ going to assign, say, to the concrete proper part of p that consists of the fusion of a and b? It will have to be zero, if we are to avoid saying that p has a degree of existence greater than 1, yet a and b are in direct contact with one another and their fusion seems like a perfectly good, and important, part of p. (This is quite unlike the case of the notional fusion of the port shear and starboard garboard strakes.)" There is no problem here for my view. Rather, the problem is an incoherence in the set of intuitions which says *both* that $\mu(a) = \mu(b) = \mu(c) = \mu(d) = \frac{1}{4}$ *and* that the fusion of a and b is a perfectly good, and important, part of p. The total measure of importance of *all* the notional parts of p must be 1. (When I talk of the measure of importance of a part, I mean the value which μ assigns to the set containing only that part—see note 48.) Parts of no importance at all get measure 0. The rest—the not-totally-unimportant parts—get measures which sum to 1. So if you think that there are more than four important parts (you think a, b, c, d *and* the fusion of a and b are all important, to some degree) then you simply should not divide the total measure of importance of 1 equally among only *four* of these parts. (Compare dividing up a cake at the birthday party of twins Bill and Ben: you give Bill and Ben half each, and then think that all the guests should have some too—but there is none left! This does not show that the cake was too small: it shows that you divided it in the wrong way.) You should say *for example* that each of a, b, c, and d has measure $\frac{8}{49}$, that each of *ab* (the notional fusion of a and b), *bc*, and *cd* has measure $\frac{4}{49}$ (while scattered fusions such as *ac* and *ad* have measure 0), that each of *abc* and *bcd* has measure $\frac{2}{49}$ (while *acd* and *abd* have measure 0), and that *abcd* has measure $\frac{1}{49}$. This renders *ab* half as important as a and b, *abc* half as important as *ab*, and so on (but of course, if your intuition concerning the relative importance of these parts is different, you can choose the numbers differently); it renders scattered fusions totally unimportant; and the measures of all not-totally-unimportant parts sum to 1. Of course you might not agree at all with this particular choice of numbers, but the general point remains: divide the total measure of importance of 1 among the parts which you think are not-totally-unimportant so that each of them gets some; do not first divide it among only *some* of these parts, and then find there is none left for other parts which you think are not-totally-

Now, in fact the proposal just described for gunk also works for the infinite nongunk case. Thus our final principle of concrete mereology is (4) above, except that S_x is the set of *all* notional parts of x:

$$\text{C3: } D_{(w,t)}(x) = \int_{S_x} f^x_{(w,t)} d\mu$$

Although there are further subtleties to be discussed in later sections, the basics of my proposal are now on the table. Our quantifiers $\exists$ and $\forall$ range over D, and they are classical. Our existence predicate E has $D_{(w,t)}$ as its extension at (w,t), and this is a fuzzy subset of D. We say that an object exists *simpliciter* if it is in the range of the quantifier $\exists$, that is, if it is in D, and that it concretely or actually exists at (w,t) to the extent that it is a member of the domain of (w,t), that is, $D_{(w,t)}$. Concrete existence can be a vague matter, but existence simpliciter is always a nonvague, all-or-nothing matter. For *any* set of things that exist simpliciter, there exists simpliciter a notional fusion of that set; however, in general these fusions *exist to degree 0 at the actual world.* A notional fusion exists at a world-time to the extent to which its notional parts are *concrete* parts of it at that world-time. This view offers a double advantage: it makes perfectly clear sense, *and* it satisfies widespread intuitions about composition and existence. It makes sense because the quantifiers and the background set D are classical: vague existence is not mysterious at all—no more so than the notion of a fuzzy subset. On the other hand, the view satisfies our intuitions, because it allows (for example) that the planks that I have acquired *gradually* come to compose a boat, which comes *gradually* into existence. Of course, the notional fusion of the planks that I have acquired exists all along as a possible object—but it only comes gradually to exist as part of the furniture of the actual world, alongside the shed in which I am building it, myself, and all the other actual, concrete inhabitants of our world. It is precisely this double picture of existence that allows my proposal to *both* make perfectly clear sense, *and* satisfy widespread intuitions about vague composition and existence.

At this point, the reader might be thinking that I am not opposed to the orthodox view at all: I am in fact an advocate of unrestricted composition (and hence my view does *not* accommodate the common intuitions about composition with which—as noted at the outset—unrestricted composition conflicts). It is true that I advocate unrestricted composition (I think that arbitrary notional fusions exist

---

unimportant. (This talk of us choosing the numbers we like is not meant to imply that importance measures—and hence degrees of concrete existence of objects—are subjective matters. We can all say what we think is the most reasonable importance measure in a given case—just as we can all say how many marbles we think the jar contains—but for all that, there may (or may not) just *be* one *correct* measure.)

simpliciter); but it is false that I am not opposed to the orthodox view (and that my view does not accommodate the common intuitions with which unrestricted composition conflicts). Let me explain. What conflicts with the common intuitions about composition is universalism *end of story*: the claim that not only tables and chairs exist, but also arbitrary fusions thereof (and so on), and that is all there is to it. Now in order *not* to be a universalist *end of story* one can either keep one's story the same length and make it nonuniversalist, or one can retain the universalism but not make it the end of the story—that is, one can add to the story. The problem with the first option—with having just one notion of existence and one notion of composition (like the universalist *end of story*), but *not* being a universalist about the existence of fusions—is that in order to satisfy our intuitions, we need a vague restriction on the existence of fusions (existence *simpliciter*, that is—for ex hypothesi, that is the only notion of existence available), and no one has been able to see how to make sense of this idea (myself included). Thus, I take the second option: I start with universalism about the existence simpliciter of notional fusions, but I then add to this a view about concrete composition and concrete existence. Thus, according to my view, arbitrary notional fusions do exist simpliciter, *but they do not also exist in the further sense in which tables and chairs exist*: notional fusions are there in the domain of all possible objects, but (in general) they do not exist in our world—*they are not present in our world alongside the tables and chairs and so on which do inhabit our world.* This seems to me to capture entirely the intuition that (in general) arbitrary fusions do not exist: *the intuition being that, unlike tables and chairs, arbitrary fusions are not part of the furniture of our world—that is, would not appear in a complete inventory of the contents of our world.* Thus I am a universalist, but I am not a universalist *end of story*—and this difference is crucial. Universalism *end of story* flouts intuition; nonuniversalism *end of story* is obscure if not nonsensical; only the present combination of universalism *plus* a further story about vague restrictions on the *concrete* existence of notional fusions offers the double advantage of *both* making perfect sense *and* being intuitively satisfying.

IV. PERSISTENCE

Consider all the objects which are (according to common sense) parts of my boat at t_1; let a be one of these objects (a plank), and let A be the notional fusion of all these objects. Suppose that a is disconnected or destroyed in such a way that it becomes a degree-0 concrete part of A, and suppose that at time t_2, the missing part of my boat is replaced by b. Consider all the objects which are parts of my boat at t_2; let B be the notional fusion of all these objects. At t_2, A no longer fully concretely exists, whereas B now does (and we may suppose that at t_1, B did not fully concretely exist, whereas A did). Thus, if we want to say that my boat—the same one as before—still fully exists, after a has been replaced by b, then we cannot identify my boat with A (or with B).

More generally, if we wish to avoid *mereological essentialism* for ordinary persisting objects such as houses and boats,[51] then we cannot identify such objects with notional fusions. But as things stand, we have nothing else in our picture with which to identify them. We have allowed that a fusion of planks can fade into and out of (actual, concrete) existence as its notional parts become concrete parts to a greater or lesser degree. But there is another way in which objects can vary over time: an object can be composed of different material at different times—it can have different parts at different times. What we now need to do is make room for this second sort of variation with time. There are two different ways of doing this: a way which involves distinguishing between material objects and the matter which constitutes them (persistence and change involve an object being differently *constituted* at different times); and a way which involves distinguishing between four-dimensional objects and their temporal parts (persistence and change involve an object having different *temporal parts* at different times).[52] I shall consider these ways in turn.

IV.1. Matter versus Material Objects. The first way of proceeding retains the above account of notional mereology as it is, and adds something new. The thought is this: so far, what we have in D are *items of matter*; but we also need to make room for *material objects.* So we add a new domain D' of material objects. D' contains every function from the set of all world-time pairs (w,t), with $w \in W$ and $t \in T_w$, to D. The function associated (or identified) with a material object tells us which item of matter—that is, which member of D—constitutes that material object at each time in each world.[53]

Our model theory undergoes the following changes: (i) ν now assigns members of $D \cup D'$ (not D) to variables, but V still assigns n-tuples of members of D (not $D \cup D'$) to predicates (at world-times). We replace the clause:

---

[51] That is, the claim that such objects have their parts essentially, and hence cannot survive the gain or loss of parts; see Roderick M. Chisholm, *Person and Object: A Metaphysical Study* (London: George Allen and Unwin, 1976), chapter III; and Simons, p. 177 (Simons uses an alternative terminology of "mereologically variable" and "mereologically constant" objects).

[52] For general discussions of these two sorts of view, see, for example, David Wiggins, "On Being in the Same Place at the Same Time," *Philosophical Review*, LXXVII (1968): 90–95; and Mark Johnston, "Constitution Is Not Identity," *Mind*, CI (1992): 89–105 (on the first sort of view); and Sider (on the second sort).

[53] I take the functions in D to be total functions, each of which assigns the null fusion to world-times at which the material object associated with the function does not exist. We stipulate that the null fusion exists to degree 0 at every world-time. Those who find the null fusion uncongenial could instead consider partial functions, each of which assigns *nothing* to world-times at which the material object associated with the function does not exist. This would require certain alterations in what follows, but none that involves any substantial difficulty.

$$V^M_v(Px_1,\ldots,x_n,w,t) = V^M(P,w,t)(v(x_1),\ldots,v(x_n))$$

with the clause:

$$V^M_v(Px_1,\ldots,x_n,w,t) = V^M(P,w,t)(m_{(w,t)}(v(x_1)),\ldots,m_{(w,t)}(v(x_n)))$$

where:

$$m_{(w,t)}(v(x)) = \begin{cases} v(x) \text{ if } v(x) \in D \\ f(w,t) \text{ if } v(x) = f \in D' \end{cases}$$

Thus, the truth (at a given world-time) of something said of a *material object* is determined by, as it were, passing the statement on to the *matter* which constitutes the object at that world-time, and seeing how true the statement is of that matter (at that world-time).[54] (ii) Q now assigns subsets of $D \cup D'$ (not D) to each (w,t). For $f \in D'$, we stipulate that $D_{(w,t)}f = D_{(w,t)}f(w,t)$, that is, for a material object f in D', concrete existence is governed by the principle that the degree to which f concretely exists at (w,t) is the same as the degree to which its constituting matter at (w,t), that is, $f(w,t)$, concretely exists at (w,t). So the account of concrete existence in section III is taken to apply directly only to items of matter (members of D); for material objects (members of D'), we now have an indirect account, parasitic on the direct one.[55]

Returning to the case of my boat, we find the case is problematic no more. My boat is not a mereologically static item of matter—a member of D. Rather, it is a material object—a member of D'. So the degree to which my boat exists at a time (in the actual world) is the degree to which its constituting matter exists at that time, where the latter is determined in accordance with the account presented in section III. At t_1, my boat is constituted by A, which concretely exists to degree 1 at t_1, so my boat concretely exists to degree 1 at t_1. (No matter that B does not concretely exist to degree 1 at t_1, because at t_1, my boat is not constituted by B.) At t_2, my boat is constituted by B, which concretely exists to degree 1 at t_2, so my boat concretely exists to degree 1 at t_2. (No matter that A does not concretely exist to degree 1 at t_2, because at t_2, my boat is no longer constituted by A.) Thus my boat—one and the same boat—exists to degree 1 at t_1 and at t_2; the trick is that this one boat is constituted by different matter at t_1 and at t_2.[56]

---

[54] With the exception, of course, of any statements involving the constitution function f itself (for example, 'This material object is constituted by that item of matter at time t in world w').

[55] Note that on this approach, corresponding to each item of matter a in D, there is a material object which is composed of that matter at all world-times: it is the constant function f_a in D' which assigns a to all pairs (w,t).

IV.2. Temporal Parts. The other way to proceed involves making notional parthood a world-time-indexed relation, rather than an absolute one.[57] We abandon the absolute relation $\sqsubset$ in favor of the indexed relation $\sqsubset_{(w,t)}$, where $x \sqsubset_{(w,t)} y$ means that x is a notional proper part of y *at time t in world w*. Note that although this relation is now (like the concrete parthood relation) nonabsolute, it is still (*unlike* the concrete parthood relation) *non*vague. The definitions, and first four principles, of section II undergo the following modifications:

- proper-or-improper part at (w,t): $x \sqsubseteq_{(w,t)} y := x \sqsubset_{(w,t)} y \vee x = y$
- overlap at (w,t): $x \boxdot_{(w,t)} y := \exists z(z \sqsubseteq_{(w,t)} x \wedge z \sqsubseteq_{(w,t)} y)$
- disjointness at (w,t): $x \boxtimes_{(w,t)} y := \neg x \boxdot_{(w,t)} y$
- atom or simple at (w,t): $A_{(w,t)}x := \neg\exists y(y \sqsubset_{(w,t)} x)$
- fusion at (w,t): $xF_{(w,t)}S$ (that is, x is a fusion-at-(w,t) of a set S) just in case:
    1. everything in S is part-at-(w,t) of x, that is, $\forall y \in S(y \sqsubseteq_{(w,t)} x)$; and
    2. every part-at-(w,t) of x overlaps-at-(w,t) something in S, that is, $\forall y(y \sqsubseteq_{(w,t)} x \rightarrow \exists z \in S(y \boxdot_{(w,t)} z))$.[58]

N1′: $\neg x \sqsubset_{(w,t)} x$
N2′: $x \sqsubset_{(w,t)} y \wedge y \sqsubset_{(w,t)} z \rightarrow x \sqsubset_{(w,t)} z$
N3′: $x \sqsubset_{(w,t)} y \rightarrow \exists z(z \sqsubset_{(w,t)} y \wedge z \boxtimes_{(w,t)} x)$
N4′: $\neg A_{(w,t)}x \wedge \neg A_{(w,t)}y \wedge \forall z, w', t'(z \sqsubset_{(w',t')} x \leftrightarrow z \sqsubset_{(w',t')} y) \rightarrow x = y$.

Let F be the set of all (total) functions from the set of all pairs (w,t) (with $w \in W$ and $t \in T_w$) to the power set of D; that is, each function in F assigns a subset of D to each world-time pair. Our principle governing the existence of fusions is then:

> N5′: For every f in F, there exists exactly one x such that for every world-time (w,t), x is a fusion at (w,t) of $f(w,t)$.

The idea is this: for each world-time, pick a bunch of objects; there then exists an object which has as its parts at each world-time exactly the things which you just associated with that world-time. Thus, for example, there is an object which consists of my now-parts now, my tomorrow-parts tomorrow, and so on (that object is me), and there are also gerrymandered objects such as one which consists of the Eiffel Tower's now-parts now and the Opera House's tomorrow-parts tomorrow.

---

[56] Other puzzle cases are handled in a similar way. For example, suppose that after t_2 I dismantle my boat and assemble all the pieces at t_3 into a sculpture that does not float and looks like a random hunk of junk, but is structurally sound. At t_3, B once again concretely exists to degree 1—but it would seem that my boat does not exist at all (supposing I have not rebuilt it from other parts). No problem: B is not the constitutor of my boat at t_3; the sculpture is made of the right stuff to be my boat, but it is not made in the right way.

[57] Sider, p. 57, discusses temporally relativized parthood.

[58] This definition follows Sider, p. 58.

Turning now to concrete mereology,.our linking principle becomes:

$$\text{NC}': \neg(x \sqsubset_{(w,t)} y) \rightarrow \lhd_{(w,t)}(x,y) = 0.$$

This says that if x is *not* a notional part of y *at a given world-time,* then x is a concrete part of y *to degree 0* at that world-time. C1 and C2 remain as above. As for C3, which governs the relationship between concrete parthood and concrete existence, we retain our earlier principle with just one change: rather than the set of all notional parts *simpliciter* of x, S_x is now the set of all notional parts *at* (w,t) of x:

$$\text{C3}': D_{(w,t)}(x) = \int_{S_x} f^{x}_{(w,t)} d\mu$$

Thus the degree to which an object x concretely exists at (w,t) is the degree to which its notional parts at (w,t) are concrete parts of it at (w,t).

Note that on this second approach we make no distinction between matter and material objects: we just have D, not D and D'. The idea is that I, for example, am the object in D which is the notional fusion of my notional now-parts now, my notional yesterday-parts yesterday, and no parts at all at times before I was born, and that I concretely exist now to the degree to which my notional now-parts are now concrete parts of me, and I exist to degree 0 before I was born (because the fusion of no parts is the null object, which exists to degree 0 at every world-time). Returning to the case of my boat: there are functions in F which assign to each time in the actual world all the objects which are (according to common sense) parts of my boat at that time, and for each such function f there is an object in D which at every world-time (w,t) is the fusion of $f(w,t)$; my boat is one of these latter objects. At t_1 it consists of all the objects which are (according to common sense) parts of my boat at t_1, and as these parts are all degree-1 concrete parts of my boat at this time, it concretely exists to degree 1. At t_2 it consists of a different bunch of objects (a has been replaced by b) and as these parts are all degree-1 concrete parts of my boat at this later time, it again concretely exists to degree 1. Thus my boat—one and the same boat—exists to degree 1 at t_1 and at t_2; the trick this time is that this one boat has different notional parts at t_1 and at t_2.

The temporal parts approach to persistence just outlined faces a subtlety which is worth discussing before we move on. Consider a particular tree, which—on the temporal parts approach—consists of (among other things) the fusion-at-t of certain tree-parts at t and the fusion-at-t' of certain tree-parts at t' (where t' is after t); consider also a particular lamp, which consists of (among other things) the fusion-at-t of certain lamp-parts at t and the fusion-at-t' of certain lamp-parts at t'; and consider finally a gerrymandered object X which consists of (among other things) the tree-parts at t and the lamp-parts at t'.

Suppose the tree is quite intact at *t* and the lamp quite intact at *t*′, so that the tree has degree-1 concrete existence at *t* and the lamp has degree-1 concrete existence at *t*′. In this case, the gerrymandered object *X* has degree-1 concrete existence at *t* and at *t*′ also. Now you might say that this is no problem: while we do not want to admit the degree-1 concrete existence at *t* of something which consists at *t* of the tree-parts at *t* and the lamp-parts at *t*, the object *X* is quite different, and entirely unobjectionable: all it takes for this object concretely to exist at *t* is for the tree concretely to exist at *t*; the latter is unproblematically the case; hence *X*'s concrete existence at *t* is unproblematic. Alternatively, one might wish to deny the concrete existence of *X* at *t*. If so, all is not lost on the present framework: one can have what one wants by imposing the principle that (for any *x* and *y* and any world-time (*w*,*t*)) if *x* is a concrete part of *y* at (*w*,*t*), *then unless x is moved between t and t*′ (or disturbed in some other way relevant to concrete parthood), *x* is a concrete part of *y* at (*w*,*t*′). With this principle in place, we find that *X* concretely exists to degree 0 at *t* (even though the tree concretely exists to degree 1 at *t*): the tree-parts at *t* are *not* concrete parts of *X* at *t*, because if they were they would have to be concrete parts of *X* at *t*′ (given our principle, and assuming that the tree does not move between *t* and *t*′), but they are *not* concrete parts of *X* at *t*′, because (and here I appeal to the linking principle NC′) they are not even notional parts of *X* at *t*′ (at *t*′, *X* consists of the lamp-parts at *t*′).

Although formally quite different, both approaches to persistence just outlined have an important feature in common. The degree to which a persisting object concretely exists at a time *t* is the same as the degree to which its temporal part/constitutor at *t* concretely exists at *t*—where the *latter* degree is determined as in section III. Thus, when it comes to the concrete existence of an ordinary persisting object *O* at a time *t*, there are *two* variables: first there is the question of the degree of concrete existence of different notional fusions at *t*; second, there is the question of which of these notional fusions is *O*'s temporal part/constitutor at *t*.[59]

---

[59] Suppose that I am building a boat, and I have all my timber scattered in the shed. Let *X* be the notional fusion of the timbers (either simpliciter, if we are working in the first framework, or at the particular time in question in the actual world, if we are working in the second framework). Each timber is a degree-0 *concrete* part of *X* at the time in question, and so the degree of concrete existence of *X* at the time in question is 0. The degree of existence of my boat at this time is, of course, also 0. But here we face a subtle question. Does my boat exist to degree 0 at this time because it *has no* temporal part/constitutor at this time (that is, its temporal part/constitutor at this time is the null fusion), or does it exist to degree 0 at this time because its temporal part/constitutor at this time is *X*, which exists to degree 0 at this time? That is, at a time before it has been built, but after its component planks have been collected in my shed, does my boat have no temporal part/constitutor, or is its temporal part/constitutor the fusion at this time of the component planks? I find the former option more congenial, but nothing hangs on this.

V. COUNTING VAGUE EXISTENTS

One of the major problems with vague composition and existence is supposed to be that they lead to indeterminacy or vagueness about *how many things there are in the world*—and this is supposed to be impossible/obscure/nonsensical. This sort of worry is central to Lewder's argument for unrestricted composition (as we shall see in section VI), and—referring to Lewis's part of that argument—Rosen and Dorr write that a vague principle of composition "is objectionable if construed as a serious theoretical claim. It seems to entail the deeply obscure doctrine that it is a vague matter how many things there are" (*op. cit.*, p. 31).

There is certainly an issue about what to say in answer to the question 'How many things are there in the world?', when things can have vague existence—but it is in no way a deeply puzzling or problematic issue.[60] Suppose that we are constructing desks to fill a school hall. We have made a certain number, and a few more are half built. How many concretely existing desks are there in the room? It is not clear how to answer this question. But note that the problem is simply an instance of a general problem which arises when we have fuzzy sets. Suppose there are five definitely bald men in the room, and a few more who are bald to various intermediate degrees. How many bald men are there in the room? What we have here is, in essence, exactly the same problem as we have in the case of the school desks. So our counting problem is a problem that arises for anyone who countenances fuzzy sets: it is *not* a special problem to do with vague *existence.*

What are we to say in answer to the problem? One option is to introduce some reasonable method for counting fuzzy sets. One thought which springs to mind is that if there were three definitely bald men and one half-bald man, we might say that the number of bald men is three and a half. Following this thought, we could say that when counting bald men, a degree-*n* bald man adds *n* to the count of bald men, so if there are three degree-1 bald men, one degree-0.3 bald man, one degree-0.2 bald man, and the rest are degree-0 bald men, then the number of bald men is 3.5. Now we might say this, but when we spell out the thought, as I just have, I think we see that it is best regarded as a *joke.* We might then look for some *other* reasonable method for counting fuzzy sets. Alternatively, we might refuse to introduce any such method, saying simply that we *cannot* count the bald men (for example). I believe the latter option is the correct one. This option faces two apparent problems, which I shall now address.

---

[60] To clarify: there is no deeply puzzling or problematic issue here when we approach vague existence as I have above—that is, if we think of '(concretely) exists' as a predicate. This is not to say that there is no deeply puzzling or problematic issue in this area for someone who thinks that the unrestricted existential quantifier $\exists$ is vague. Such a person really would have a big problem here—and as a referee pointed out, "people arguing against the possibility of counting vague existents have in mind a purely Quinean (that is, logical) notion of existence." So I do not mean to suggest that I am solving this big problem. Rather, the point is that my approach to vague existence never runs into this big problem in the first place. This is a very good thing, because the problem is (as far as I can see) insuperable—that is, I *agree* with those who deny that $\exists$ can be vague.

First there is the problem of expressive power. If I cannot count the membership of fuzzy sets, then there are huge expanses of reality which I am simply powerless to describe (for example, I cannot say how many bald men there are in the room); this is a bad thing; so we should introduce methods for describing these parts of reality. But the problem is not as bad as it may seem. First, we can *precisify.* We can count any crisp set, so we can count the things which are bald to degree 1, the things which are bald to a degree greater than 0.3, the things which are between 0.2 and 0.4 bald, and so on. Second, consider the numerical sentences of first-order logic:

(1) There is exactly one P: $\exists x(Px \wedge \forall w(Pw \rightarrow w = x))$
(2) There are exactly two Ps: $\exists x \exists y(Px \wedge Py \wedge x \neq y \wedge \forall w(Pw \rightarrow w = x \vee w = y))$
(3) There are exactly three Ps: $\exists x \exists y \exists z(Px \wedge Py \wedge Pz \wedge x \neq y \wedge x \neq z \wedge y \neq z \wedge \forall w(Pw \rightarrow w = x \vee w = y \vee w = z))$
$\vdots$

In a classical context, for any interpreted predicate P (that is, for any set P), either none of these sentences will be true, or exactly one will be true—and if the nth sentence is true, there are exactly n Ps. When P is a fuzzy set, this breaks down; but nevertheless, we are not left with nothing to say: we can calculate the truth value of each of these numerical sentences, and we can then say that the degree of truth of the claim that there are exactly n Ps is the degree of truth of sentence n.[61] Thus, in response to the problem of expressive power: we can count all the degree-n bald men, for each n; and we can give the degree of truth of each claim 'there are exactly n men'; and once we have done these things, it seems to me we have given as much information as anyone could reasonably want—we do not need to then go on to give the number of bald men *simpliciter.*

The second problem is a Quinean one. You might think that the problem with not being able to count the concretely existing boats, the bald men, and so on, is that these things will then get mixed up with one another—they will not be able to be individuated properly—and then, following Quine, you might think that this is such a serious problem that we cannot seriously speak of such things at all. But in the present case we run into no such problems. There is no problem of individuation, no confusion about identity: the impossible-to-count concretely existing boats do not get mixed up with one another, because when we try to count them, we are not trying to count objects *simpliciter*, we are trying to count a fuzzy subset of a background set of objects (the overall domain D, which is a regular, classical set), and identity and individuation are clear and precise matters in this background set. We cannot count the concretely existing boats *not* because each of them is (as it were) a fuzzy dot which blurs into other dots in such a way that we cannot get a handle on how many dots there are—this is indeed deeply obscure, if not outright nonsense; rather each is (continuing the metaphor) a precise dot, quite distinct from all other dots and in no danger of being confused with them—but the borders of the *set* of things we are counting is fuzzy, so that some of these precise dots lie in the penumbral region of the set, and hence we do not know whether or not to include them in our count.

---

[61] Compare the use of such numerical sentences in the theory of indeterminate identity developed in Terence Parsons, *Indeterminate Identity: Metaphysics and Semantics* (New York: Oxford, 2000).

In sum, the response to the counting problem is that there is no answer to the question 'How many Ps?' when P is a fuzzy subset—although the precisified questions 'How many degree-n Ps?', 'How many degree-n-or-less Ps?', and so on, all have answers, and the claims 'There is exactly one P', 'There are exactly two Ps', and so on, all have unique fuzzy truth values (so even though there is no uniquely correct answer to the simple question 'How many Ps?', some possible answers are better—that is, more true—than others). Of course, the *problem* was supposed to be precisely that there is no answer to the question 'How many Ps?' when P is a fuzzy subset: so to be quite explicit, the response is that this is not a problem at all.

VI. LEWDER'S ARGUMENT

I have set my views in opposition to Davore Lewder's "orthodox argument," which is a rough compilation of two arguments in the literature. I shall now respond to these two arguments.

Lewis has a famous argument for unrestricted composition:[62]

(1) A restriction on composition must be vague or nonvague.

(2) "To restrict composition in accordance with our intuitions would require a vague restriction." So a *nonvague* restriction would fail to satisfy the intuitions motivating restriction in the first place, and hence "would be gratuitous."

(3) "But if composition obeys a *vague* restriction, then it must sometimes be a vague matter whether composition takes place or not. And that is impossible." The reason it is impossible is that:

---

[62] Quotations are from Lewis, *Plurality*, pp. 212–13.

(a) "The only intelligible account of vagueness locates it in our thought and language...Vagueness is semantic indecision"; and

(b) "The question whether composition takes place in a given case... can be stated in a part of language where nothing is vague. Therefore it cannot have a vague answer."

(4) So we cannot have a vague restriction on composition, and we have no use for a nonvague one. Conclusion: composition is unrestricted.

An obvious line of objection to this argument is this: Why should we accept that the predicate 'is part of' is not vague (this predicate being in the part of language needed to ask the question whether composition takes place in a given case—see step 3b)? For it seems very natural to say, for example, that the hair that Tibbles is in the process of shedding is only *sort of* part of Tibbles. Enter Sider, who argues that it can never be a vague matter whether or not composition takes place, "for if it were vague whether a certain class had a fusion then it would be vague how many concrete objects exist" (*op. cit.*, pp. 126–27). In the terms of my presentation of Lewis's argument above, Sider is defending step 3, but in a new way—one which avoids the need for a direct argument that the predicate 'is part of' is nonvague. Sider argues as follows. Let the *concrete* objects be those objects which are *not* sets, classes, numbers, properties, and so on (add your favorite "abstract" objects to the list).

> Suppose now for *reductio* that...it can be vague whether a given class has a fusion. In such a case, imagine counting all the concrete objects in the world. One would need to include all the objects in the class in question, but it would be indeterminate whether to include another entity: the fusion of the class. Now surely if [the claim that it can never be vague whether a given class has a fusion] can be violated, then it could be violated in...a world with only finitely many concrete objects. That would mean that some *numerical sentence*—a sentence asserting that there are exactly n concrete objects, for some finite n—would be indeterminate. But numerical sentences need contain only logical terms and the predicate 'C' for concreteness (a numerical sentence for $n = 2$ is: $\exists x \exists y[Cx \,\&\, Cy \,\&\, x \neq y \,\&\, \forall z(Cz \rightarrow [x = z \vee y = z])]$) (*op. cit.*, p. 127).

Now given the linguistic theory of vagueness, a numerical sentence can only be indeterminate if one or more of its components has multiple admissible interpretations. Sider argues, however, that all components are precise: apart from the logical vocabulary, which is precise,[63] there is only the predicate '*C*', and it is precise because "it was defined by a list of predicates for fundamental ontological kinds that do not admit of borderline cases" (*op. cit.*, p. 127). The conclusion is that it *cannot* be vague whether a certain class has a fusion.

---

[63] Although I agree with Sider that the logical connectives are precise, Sider begs the question when he says that it "is overwhelmingly plausible that the boolean operators lack precisifications" (p. 128). Of course the *Boolean* operators lack precisifications: but reading the logical operators in Sider's numerical sentences *as Boolean operators* is *already* choosing one way of precisifying them. In general, the different precisifications of the logical operators correspond to the different possible systems of semantics for the logical connectives: for example, in the case of the negation

Although I have employed a nonlinguistic theory of vagueness in this paper, *this is not the source of my disagreement with Lewis and Sider.*[64] Rather, my disagreement is as follows. First, Lewis: obviously, I think that the question whether composition takes place in a given case can*not* be stated in a part of language where nothing is vague—for 'part' (that is, '*concrete* part') is vague.[65] Second, Sider: Sider says that if composition were vague (in a finite world), then some numerical sentence would be indeterminate. Is this true? Suppose that the numerical sentences involve *Sider's* predicate '*C*' for concreteness. In this case, it does *not* follow from my views about vague composition that some numerical sentence will be indeterminate: the overall domain *D* will be precisely cleaved into two crisp subsets, containing the concrete (in Sider's sense) and the nonconcrete objects respectively; the fact that (for example) some table only concretely exists—in *my* sense of 'concretely exists'—to degree 0.5 at the present time in the actual world does not make it any less than *definitely* a concrete object in *Sider's* sense of 'concrete'. Suppose instead that the numerical sentences involve *my* predicate '*E*' for concrete existence. In this case it *does* follow from my views about vague composition that some numerical sentence will be indeterminate (at some times in some worlds); but as shown in foregoing sections, this is perfectly possible—and it is even possible on a linguistic theory of vagueness.[66]

VII. FURTHER ISSUES

In this section I discuss four further issues connected with the views presented in this paper. Some of these issues merit a more extended treatment in future work, but here I can only touch on them briefly.

VII.1. Ambiguity. Suppose that I am building a boat: currently it is about half built, and I have not yet bought all the timber required to finish it. In this case when I speak of 'the boat I am building', it

---

operator, we have a three valued Boolean reading, a three valued de Morgan reading, a two valued classical reading, and so on.

[64] It had better not be: for recall that I noted at the outset that everything that I do in this paper could also be done within the framework of a linguistic theory of vagueness (in particular, within the degree-theoretic form of supervaluationism).

[65] How could 'part' be vague on a linguistic theory of vagueness? What would be its precisifications? Well, each precisification would involve a precise answer to the question as to how much contact (or in general, proximity) and adhesion is required between objects in order for them to be concrete parts of their notional fusion.

[66] In the latter case each precisification would involve a precise answer to the question as to how much of an object has to be concretely assembled in order for the object to exist concretely.

seems that there is no clear referent for this phrase. There are many boats in many possible worlds, each of which may, for all I know, turn out to be the boat I end up building. This in itself is not a great problem: we may suppose that there is one particular boat that I will, as a matter of fact, end up building, and the fact that I do not yet *know* which boat it is need have no effect on the *semantics* of the situation. But the problem gets worse: for suppose that as a matter of fact I shall never finish my boat. Nevertheless, it still seems to make good sense to speak of 'the boat I am building' (or later, of 'the boat I was building before I gave up'). Here we may use supervaluations. When I use a term, such as 'the boat I am building', which has multiple admissible referents (in the present case, each possible completion of my boat is an admissible referent)—or when I use a predicate with multiple admissible extensions—we have a notion of an *admissible interpretation* of my utterance. Each admissible interpretation is a model of the sort discussed above; in each admissible interpretation, each initially polysemic term is assigned a unique referent/extension (in an acceptable way: this is what makes the interpretation an *admissible* one).[67] Now, suppose I make an utterance involving one or more polysemic terms. This utterance has a particular degree of truth in each admissible interpretation. Overall, it is true simpliciter, or supertrue, if it is true on *every* admissible interpretation. Now, we are working in a degree-theoretic environment, so if $\{\mathcal{M}_i\}$ is the set of all admissible interpretations of my utterance, and f_i is the truth value of my utterance on $\mathcal{M}_i$, then the supertruth value of my utterance with respect to the supervaluation $\{\mathcal{M}_i\}$ is $\bigwedge\{f_i\}$ (the universal quantifier in 'supertruth is truth in *every* admissible interpretation' thus receiving its standard treatment as generalized conjunction). The case where my utterance contains no polysemic terms becomes the special case where the supervaluation has one member, and in this case the supertruth value of my utterance is the same as its truth value on this interpretation.

VII.2. Quasidentity. Lewis[68] discusses objects which are not identical, but which almost are, in the sense that they almost entirely overlap, and he argues that sometimes when we individuate objects, we are not concerned with strict identity, but with overlap. In the present framework, identity is an entirely classical, nonvague matter—but as well as nonvague identity, we also have a vague relation: (concrete) overlap. We thus have the resources to follow Lewis and talk *loosely* about degrees of 'identity'—we must simply realize that when we talk this way, we are really using 'identical' to mean overlapping. Lewis thinks that total overlap coincides with (strict) identity. Is this the case in the present framework? Certainly two nonidentical objects can entirely concretely overlap at some world-time. But does total overlap at every world-time imply (strict) identity? Maybe as a matter of fact, but not as a matter of principle: for recall the discussion of extensionality in section III, in which I argued that we should *not* adopt as a fundamental principle of concrete mereology the claim that if two objects have the same concrete parts to the same degrees at every world-time, then they are identical.

---

[67] See Kit Fine, "Vagueness, Truth, and Logic," in Rosanna Keefe and Peter Smith, eds., *Vagueness: A Reader* (Cambridge: MIT, 1997), pp. 119–50.

[68] "Many, but Almost One," in John Bacon, Keith Campbell, and Lloyd Reinhardt, eds., *Ontology, Causality, and Mind: Essays in Honour of D.M. Armstrong* (New York: Cambridge, 1993), pp. 23–38.

VII.3. Essential Parts. We could define the *essential* parts of x as the things which are positive-degree concrete parts of x at every world-time at which x exists to degree 1 (or, more stringently, at every world-time at which x exists to any degree greater than 0). Thus, an object cannot fully exist (or, on the stronger definition, exist at all) without its essential parts. Note that this provides an instant tie-in with the notion of a part's degree of importance, introduced in section III: essential parts cannot have zero importance (and on the stronger definition, the set of all essential parts of an object has degree of importance 1).

VII.4. Contact and Adhesion. I have not said much about *what makes it the case* that some notional part of a notional fusion is a *concrete* part of that fusion (at some world-time). I assume that concrete parthood often has a lot to do with contact and adhesion; however, I am not in a position here to offer a *theory* of the relationship between concrete parthood on the one hand, and physical contact and adhesion on the other. Suppose, by way of analogy, that we are theorizing the relationship between the colors of objects and the emotions they provoke in viewers. There are facts in the offing such as that no object can be both green and red. There are two attitudes we can take to these facts. (i) We can regard them simply as facts. If, according to our theory, only red provokes emotion A and only green provokes emotion B, then as a matter of fact, no uniformly colored object will simultaneously produce both emotion A and emotion B. Now this is an interesting fact, but not a theorem of our theory, given that the facts about the relationships between the colors have the status of untheorized, background facts. (ii) We can produce a 'logic of color' and graft it onto our theory. If it is a theorem of the logic of color that no object can be both green and red, and we have as theorems of our theory that only red provokes emotion A and only green provokes emotion B, then it will follow as a theorem that no uniformly colored object can provoke both emotion A and emotion B. Returning to the case of concrete mereology, I am taking the analog of approach (i) to facts about the relationship between contact, adhesion, and concrete parthood: these facts have the status of untheorized, background facts which have implications for, but are not part of the content of, the theory offered here.[69]

VIII. CONCLUSION

In this paper, I have tried to present a view which both makes perfectly clear sense, and satisfies widespread intuitions about composition and existence. On the first point: the views presented herein appeal only to tried and tested logical techniques; there are no obscure attempts to introduce, for example, new forms of vague unrestricted quantification. On the second point: first, it seems to me that the exclusion of objects such as the fusion of the Empire State Building and the Sydney Opera House from the domain of the actual world is exclusion enough to satisfy widespread intuitions about the nonexistence of fusions—not only do we do not *need* to go on to deny that such objects are present in the overall domain of possible objects, this would in fact be a counterintuitive move (given that we think it is *possible* for the fusion of the Empire State Building and the Sydney Opera House to exist: just bring the two together in an architecturally coherent and structurally sound way); and second, the view presented here allows us to say, for example, that as we transform a pile of planks into a boat, the planks *gradually* come to compose a further object, which *gradually* comes into existence.

NICHOLAS J.J. SMITH

University of Sydney

APPENDIX: FUZZY MODAL CONSEQUENCE RELATIONS

Standard discussions of modal logic focus on the notion of *valid formula* or *tautology*. A wf α is said to be *valid in a model* $\mathcal{M}$ iff $V_v^{\mathcal{M}}(\alpha,w) = 1$ for every $w \in W$ and every assignment v based on $\mathcal{M}$. A *frame* is a pair (W,R) with W and R as above; a model $\mathcal{M} = (W,R,D,Q,V)$ is said to be *based on* the frame (W,R); and a wf α is said to be *valid on a frame* (W,R) iff it is valid in every model based on that frame. I focus instead on the notion of *valid inference* or *consequence.* We can distinguish a number of consequence relations, arising from the various combinations of the two basic parameters *local–global* and *fuzzy–classical:*

---

[69] There is a substantial literature on *mereotopology*, which seeks to integrate mereological notions such as *parthood* with topological notions such as *contact* and *boundary* (see, for example, Barry Smith, "Mereotopology: A Theory of Parts and Boundaries," *Data and Knowledge Engineering*, xx (1996): 287–303; and Roberto Casati and Varzi, *Parts and Places: The Structures of Spatial Representation* (Cambridge: MIT, 1999), and the many references therein). The sort of theory I envisage—which treats of the relationship between concrete parthood, contact, and adhesion—would need to incorporate the insights of mereotopology, but would also need to go further, in discussing the nontopological concept of *adhesion*.

- $\Gamma \models^{g\text{-}c}_{\mathcal{F}} \alpha$ (classical global consequence)
  there is no classical model $\mathcal{M}$ based on a frame in class $\mathcal{F}$, such that every member of Γ is true at every world in $\mathcal{M}$, and α is false at some world in $\mathcal{M}$.
- $\Gamma \models^{g\text{-}f}_{\mathcal{F}} \alpha$ (fuzzy global consequence)
  there is no fuzzy model $\mathcal{M}$ based on a frame in class $\mathcal{F}$, such that every member of Γ is >0.5 true at every world in $\mathcal{M}$, and α is <0.5 true at some world in $\mathcal{M}$.
- $\Gamma \models^{l\text{-}c}_{\mathcal{F}} \alpha$ (classical local consequence)
  there is no classical model $\mathcal{M}$ based on a frame in class $\mathcal{F}$, such that there is a world in $\mathcal{M}$ at which every member of Γ is true, and α is false.
- $\Gamma \models^{l\text{-}f}_{\mathcal{F}} \alpha$ (fuzzy local consequence)
  there is no fuzzy model $\mathcal{M}$ based on a frame in class $\mathcal{F}$, such that there is a world in $\mathcal{M}$ at which every member of Γ is >0.5 true, and α is <0.5 true.

It is not hard to prove that $\Gamma \models^{g\text{-}c}_{\mathcal{F}} \alpha$ iff $\Gamma \models^{g\text{-}f}_{\mathcal{F}} \alpha$, and $\Gamma \models^{l\text{-}c}_{\mathcal{F}} \alpha$ iff $\Gamma \models^{l\text{-}f}_{\mathcal{F}} \alpha$, that is, the classical and fuzzy consequence relations are identical in each of the local and global cases.[70]

We get notions of *tautology* when we set $\Gamma = \varnothing$. Of the four notions we get from the above notions of consequence by setting $\Gamma = \varnothing$, it is easy to see that the local and global classical notions coincide, and the local and global fuzzy notions coincide. Given that the classical and fuzzy consequence relations are identical in each of the local and global cases, it follows that there is just one class of tautologies for any class of frames. Thus where $\mathcal{F}$ is the set of all frames, the set of classical $\mathcal{F}$-tautologies is the set of formulas provable in K, and *so is the set of fuzzy $\mathcal{F}$-tautologies*; where $\mathcal{F}$ is the set of reflexive transitive frames, the set of classical $\mathcal{F}$-tautologies is the set of formulas provable in S4, and *so is the set of fuzzy $\mathcal{F}$-tautologies*; and so on. Thus, even when we fuzzify our modal semantics, we can retain our usual systems of modal proof theory, and the usual correspondences between frame conditions and modal axioms. This makes the fuzzy modal logic presented here very tractable.

There are stricter notions of local and global fuzzy consequence:

- $\Gamma \models^{g\text{-}f\text{-}s}_{\mathcal{F}} \alpha$ (strict fuzzy global consequence)
  there is no fuzzy model $\mathcal{M}$ based on a frame in class $\mathcal{F}$, such that

---

[70] Apply the basic idea behind the proof at pp. 193–95 of Smith, "Vagueness and Blurry Sets."

every member of Γ is 1 true at every world in $\mathcal{M}$, and α is not 1 true at some world in $\mathcal{M}$.

- $\Gamma \models^{l\text{-}f\text{-}s}_{\mathcal{F}} \alpha$ (strict fuzzy local consequence)
  there is no fuzzy model $\mathcal{M}$ based on a frame in class $\mathcal{F}$, such that there is a world in $\mathcal{M}$ at which every member of Γ is 1 true, and α is not 1 true.

These do *not* coincide with, respectively, the global and local classical notions. The reason I favor the consequence relations discussed above—the ones which coincide with the classical consequence relations—is as follows. In vague contexts, there is a natural distinction between *inference grade* statements and *assertion grade* statements. A natural response to Sorites reasoning is that we cannot simply keep taking the output of the inductive premise and feeding it back into the inductive premise at the next stage of the slide down the series—or at least we can, but our conclusions become progressively shakier. ‘Man 1 is bald’ is certainly true. ‘For any *n*, if man *n* is bald then man *n* + 1 is bald’ seems a safe enough assertion. From these it follows that man 2 is bald. Now if we feed this output back in, we get ‘man 3 is bald’; if we feed this output back in, we get ‘man 4 is bald’; and so on. But intuitively, our conclusions become more and more shaky. A natural thought is that while an output at one stage may be safe enough to assert, it may not be safe enough to serve as the start of the next stage of reasoning: it is like a terminating pass, as opposed to a pass which enables you to enroll in the next level of studies. So there seems to be, as I said, a natural distinction between *inference grade* statements and *assertion grade* statements. Now if we take a statement to be inference grade if it is strictly greater than 0.5 true, and assertion grade if it is at least 0.5 true, then we can say that a valid inference in the sense introduced above always yields at least an assertion grade conclusion, when the premises are all inference grade (in the local case, if all the premises are inference grade at a world, then the conclusion is at least assertion grade at that world; in the global case, if all the premises are inference grade at every world, then the conclusion is at least assertion grade at every world). This fits perfectly with the intuition that sorites reasoning is *valid*, even though it cannot be continued indefinitely and still yield secure results.

COMMENTS AND CRITICISM

DOORS, KEYS, AND MORAL LUCK: A REPLY TO DOMSKY*

In a recent article in this journal, Darren Domsky claims to have finally solved the problem of moral luck.[1] This problem, he says, "is like a plague. It is widespread, difficult to overcome, and tends to still show symptoms even in those who claim to be rid of it" (446). Now, finally, the plague is cured. The secret of the cure lies in the recognition that, in fact, the reasons for originally believing there was a problem are groundless. Or, to mention another of Domsky's metaphors:

> The correct key is actually no key at all. It is the realization, simply, that there is no door (447).

Is there indeed "no door," and hence no need for a key to open it? Can we safely move the puzzle of moral luck to the history of philosophy department, or does it still belong in practical philosophy proper? In this paper, I opt for the latter option. I argue that the problem of moral luck is wider than the one presented by Domsky; that he fails to refute the main arguments in favor of moral luck; that the psychological explanation he offers for the spreading of the "plague" is unconvincing; and that even if it were convincing, it would fall short of solving the philosophical problem at hand.

I. PRELIMINARIES: WHAT IS THE PROBLEM OF MORAL LUCK?

According to Domsky, the problem of moral luck results from two firm but incompatible moral intuitions:

> *Intuition 1.* Moral blameworthiness is independent of luck. Equally negligent agents are equally blameworthy, even if by luck one brings about bad outcomes and one does not.
> *Intuition 2.* Moral blameworthiness is *not* independent of luck. Negligent agents who by luck bring about bad outcomes are more blameworthy than equally negligent agents who by luck do not (445).

Let me start by noting three ways this formulation of the moral luck

---

* For helpful comments on earlier versions, I am greatly indebted to Darren Domsky, Norvin Richards, and Saul Smilansky.

[1] "There Is No Door: Finally Solving the Problem of Moral Luck," this JOURNAL, CI, 9 (September 2004): 445–64.

0022-362X/05/0208/422–36

problem unjustifiably narrows the problem at hand. First, the focus on *blameworthiness.* On the above formulation, the contradictory intuitions regarding moral luck concern the necessary conditions for blameworthiness. More precisely, they concern the validity of one such condition, that is, the condition of being-independent-of-luck. On intuition 1, a necessary condition for some agent *A* being blameworthy for some result *r* is that *A*'s bringing about *r* was not a matter of luck, while on intuition 2, no such condition exists, hence *A* might be blameworthy for *r* even if *A*'s bringing it about was a matter of luck. This means that when two agents are equally negligent and one brings about worse outcomes than the other, then, on intuition 1, they are equally blameworthy, while on intuition 2, one agent is more blameworthy than the other. In other words, on intuition 1, blameworthiness cannot extend beyond the boundaries of control, while on intuition 2 blameworthiness can so extend, hence one might be—and very often is—blameworthy for outcomes that rest beyond one's control.

The problem with this formulation of the moral law problem is that it makes no sense to speak of blameworthiness *simpliciter,* since blameworthiness is always attributed *for* some negative state of affairs, or negative behavior. Domsky says that on intuition 1, "equally negligent agents are equally blameworthy, even if by luck one brings about bad consequences and one does not." But if there are no bad consequences, what consequences *could* one be blameworthy for? Similarly with intuition 2, Domsky says that "negligent agents who by luck bring about bad outcomes are more blameworthy than equally negligent agents who by luck do not." To say that the unlucky agents are *more* blameworthy than the lucky ones assumes a fixed point of reference, that is, some unwelcome outcome, in relation to which they are more blameworthy than the lucky agents. But this is misleading. If *A* and *B* are equally negligent in their driving, they are equally blameworthy for the irresponsible endangering of human lives created by their driving. And if, because of the negligence, *A* kills some human being, while *B* kills nobody, it makes no sense to say that *A* is *more* blameworthy than *B* *for* such killing *C,* for the obvious reason that it bears no causal connection to *B*'s behavior. The assumed paradox between the above two intuitions seems to dissolve once the logic of blameworthiness is revealed: as far as the negligent behavior is concerned, then surely if two agents are equally negligent, they are equally blameworthy for their bad behavior. While as far as the outcome of the behavior is concerned, if *A* causes some unwelcome outcome while *B* does not, then *A* is blameworthy for this outcome, while *B* is not. Proponents of moral law certainly do not think otherwise, hence the above formulation fails to show what the paradox of moral law is.

In light of this analysis, it is no surprise that most writers on the moral luck problem, particularly those who first raised it in contemporary moral philosophy, did not express it through the notion of blameworthiness. For Bernard Williams, the main concern is "retrospective justification," that is, the question of how far, and in what ways, the view that an agent retrospectively takes of himself or herself may be affected by the results of his or her actions.[2] For Thomas Nagel and many others, moral luck concerns the moral judgment of people,[3] a judgment about their "moral status,"[4] "moral position,"[5] or "moral (dis)credit."[6] For still others, the problem of moral luck is much wider than the moral status of lucky versus unlucky agents. It is about "conceptions of agency and responsibility of the sort that morality requires,"[7] and it affects wide and numerous aspects of the ethical life.

In addition to clarifying the nature of the paradox under discussion, these comments raise some skepticism about Domsky's hope to have solved once and forever *the* problem of moral luck. As the problem cannot be—and usually is not—formulated in terms of blameworthiness, any solution which is based on this formulation will probably fail.

A second difficulty in Domsky's formulation of the problem of moral luck is its limitation to cases known as resultant luck, namely, "luck, good or bad, in the way things turn out."[8] Domsky is aware of other types of luck, but he believes that the problem of resultant luck "clearly dominates, not just in the literature, but in our lives" (447, note 4), and produces genuine distress, unlike the forced distress that arises in the other types of luck. I think Domsky is wrong on this point. As for the philosophical literature on moral luck, resultant luck has not received more attention (certainly not significantly more attention) than circumstantial luck (luck in the moral tests one faces).[9] And as for the dominance of resultant luck in our *lives*, I estimate that the sense of bad luck in failing a moral test which others were fortunate not to have been called on to face is far more prevalent that Domsky assumes. Just like resultant luck, we feel conflicting drives within us when it comes to luck in circumstances: on the one hand, we strongly feel that there is something unfair in this state of affairs, while on the other, we go on judging people according to what they *actually* do, not according to what they would have done had they found themselves in different circumstances. On the one hand, we believe that war criminals must suffer harsh punishment for their crimes, while, on the other, we realize that "given a context of violent conflict, given extraordinary pressures and expectations, ordinary people can act very badly indeed."[10] *Contra* Domsky, the distress produced by circumstantial luck is no less genuine than that produced by resultant luck, and limiting the discussion to the latter type of luck leaves the reader with the feeling that there is much more yet to be solved.[11]

---

[2] "Postscript," in Daniel Statman, ed., *Moral Luck* (Albany: SUNY, 1993), pp. 251–58, see p. 256; hereafter, this volume will be referred to as *ML*.

[3] "Moral Luck," in *ML*, pp. 57–71, on p. 59. Cf. Norvin Richards, "Luck and Desert," in *ML*, pp. 167–80; see also Henning Jensen, "Morality and Bad Luck," in *ML*, pp. 131–40, who talks about the "moral assessments of persons," on pp. 138–39.

[4] Judith Andre, "Nagel, Williams, and Moral Luck," in *ML*, pp. 123–30, on p. 125.

[5] Nicholas Rescher, "Moral Luck," in *ML*, pp. 141–66, on p. 154.

[6] Judith Thomson, "Morality and Bad Luck," in *ML*, pp. 195–216, on pp. 195–97.

[7] Margaret Walker, "The Virtues of Impure Agency," in *ML*, pp. 235–50, on p. 243.

[8] Daniel Statman, "Introduction," in *ML*, pp. 1–34, on p. 13.

[9] Circumstantial luck is discussed, *inter alia*, by Nagel, Richards, Walker, Thomson, and Michael Zimmerman, "Luck and Moral Responsibility," in *ML*, pp. 217–34.

Finally, even if the discussion on moral luck is limited to resultant luck, I see no reason to further limit it to cases of *negligence*, as suggested by Domsky when he says that "only negligent agents experience moral luck.... To be negligent is just to act, under conditions of uncertainty, in unjustifiably risky ways" (448). I believe that here Domsky is conflating two different types of resultant luck, one having to do with unfortunate outcomes following the taking of unjustifiable risks (negligence), and the other having to do with unfortunate outcomes ensuing from decisions taken under conditions of uncertainty. Nagel's examples of the last type include Chamberlain signing the Munich agreement, the Decembrists persuading the troops under command to revolt against the Czar, and the introduction of two people in an attempt at match-making (*op. cit.*, pp. 61–62). Unlike cases of negligence, in these cases the risk taken by the agent's decisions could not be said to have been unjustified *at the time they were taken.* Nagel is right to note that in these cases there is no possible decision, in light of what is known at the time, "which will make reproach unsuitable no matter how things turn out" (*op. cit.*, p. 62). *Contra* Domsky, then, an agent who makes a reasonable decision that, nevertheless, ends up in a catastrophe does not assess her moral position as about the same as she would have done had everything worked nicely according to plan. Nor would we, the spectators, do so. That is what is so disturbing about moral luck, especially *bad* moral luck; it strikes the righteous and the wicked alike.

---

[10] Trudy Govier, *Forgiveness and Revenge* (New York: Routledge, 2002), p. 117. Govier relies on this insight to establish the possibility and rationality of forgiveness.

[11] In fairness to Domsky, I should note that he reports that he is dealing with the other types of luck in other work in progress (447, note 4). Without having seen this work, however, it is hard to be convinced that *the* problem of moral luck has been solved.

II. THE "PSYCHOLOGICAL" SOLUTION TO THE MORAL LUCK PROBLEM

According to Domsky, the key to solving the moral luck problem lies in what I shall call the *psychological explanation* (PSE). If it sounds odd to you that a *psychological* explanation might solve a *philosophical* problem, it should. Indeed there is something suspicious going on here. Let me start by revealing what it is and by making some general remarks about methodology. Then I shall present the solution itself and explain why I find it unconvincing.

On Domsky's view, given that intuitions 1 and 2 are incompatible, the problem of moral luck is to decide which of these should be thrown away. The solution lies, he says, in realizing that intuition 2 is false, but this solution has already been advanced several times. The novelty of Domsky's argument, or so he believes, is in the complete demolition of this intuition.

> The secret to rejecting it successfully, or even to really knowing that it must be rejected, is to realize that our reasons for originally upholding it were never any good.... The solution to the problem of moral luck has eluded us simply because we have never properly understood its causes (447, 454).

The causes Domsky points to, and to which I shall turn soon, are various *biases*, particularly the selfish and the optimistic biases. Though I concede that such psychological explanations might play a role in solving philosophical puzzles, they play a much more minor role than is assumed by Domsky, and, most importantly, they are always parasitical on philosophical arguments to do the main work. A psychological explanation is required, or, at any rate, could be helpful, only when some position is shown not only to be wrong, but to be so irrational that only some fundamental cognitive bias could explain how people could ever take it seriously in the first place. In the present context, this means that exposing the psychological attraction for the belief in intuition 2 cannot by itself constitute a solution to the problem of moral luck, definitely not *the* solution to this problem.[12]

---

[12] In the last page of his paper, Domsky mentions very briefly a few philosophical arguments against intuition 2, which he further develops in a later paper entitled "Tossing the Rotten Thing Out: Eliminating Bad Reasons Not to Solve the Problem of Moral Luck," forthcoming in *Philosophy*, but in "There Is No Door," the main focus is on the psychological argument.

But even if PSE *was* admitted as an argument against moral luck, I believe it fails on its own terms, that is, it fails to explain the attraction of intuition 2. Let us, then, turn to a critical analysis of PSE itself.

The psychological key to understanding our conviction about moral luck, argues Domsky, lies in two separate cognitive biases, the selfish bias and the optimistic bias. The former is the tendency we have to select and uphold moral theories and beliefs according to how they stand to benefit us personally. Most of us are not aware of this bias and, even worse, most of us believe we are less motivated by self-interest in forming our moral convictions than the average person. The optimistic bias is the tendency to make unrealistically low estimates of the relative likelihood of experiencing unlucky outcomes. Because of this tendency, we are confident that we are somehow immune to bad luck. As for good luck, we feel it is somehow deserved. We subconsciously believe that we have a special mastery over luck, hence we feel that it is to our credit that we are luckier than our peers.

The upholding of intuition 2 is the natural child, as Domsky puts it, of these two biases (462). Because of the optimistic bias, we estimate that we are far less vulnerable to bad luck than other human beings, hence we subconsciously believe that if we found ourselves in cases like the two equally negligent drivers, we would most probably be the lucky one. And the selfish bias explains why we select the belief in intuition 2:

> We are subconsciously drawn toward moral beliefs that privilege us, and we subconsciously believe that blaming the morally unlucky much more than the morally lucky privileges us at the expense of others by making us less blameworthy than we would otherwise be and by making them more (462).

Given our psychological makeup, argues Domsky, it would be remarkable if we did *not* believe in moral luck. In any case, now that the pathetic irrationality of this makeup is revealed, especially regarding the optimistic bias, there is finally a real chance to overcome our pull towards intuition 2, and to realize that there is "no door."

Is PSE convincing? I believe not. Let us grant that in the case of principles allocating blameworthiness, just as with other moral principles, a major, subconscious reason for deciding what principles to adopt is their expected utility in terms of the agent's self-interest. Why, then, do we have stronger reasons to choose intuition 1 over intuition 2? On Domsky's view, since we believe that we are somehow immune to bad luck, we are not bothered by the fact that intuition 2 imposes more blameworthiness on unlucky agents than on lucky ones. Furthermore, since we feel that it is to our credit that we are luckier than our peers (458), we are not bothered at all by the fact that our moral status will be determined by luck. But in these respects, intuition 1 should look even more attractive, as it guarantees us *complete* immunity to luck. If only we behave reasonably, we will never be reproached for the unfortunate outcomes of our behavior, and even if we misbehave, for example, by negligent behavior, we will not be held responsible for any unfortunate results, but only for the wrongful taking of risks. If it is our own self-interest we are after, then surely intuition 1 makes the better offer and provides us with what Williams called "a solace to a sense of the world's unfairness" (*op. cit.*, p. 36). I am taking for granted here that the optimistic bias does not make us believe (even subconsciously) that we are *100% immune to bad luck.* We might be helplessly and irrationally optimistic, but we are not totally blind. As a matter of fact, we often interpret events that happen to us as "bad luck," while we attribute similar calamities that afflict others as ensuing from their mistakes or wrongful behavior.[13] True, the optimistic bias makes us believe that we are far more immune to luck than others, but assuming that we are not totally deluded, we must realize that to *some* extent we, too, are vulnerable to luck, otherwise we would invest much more money in the lottery, avoid the nuisance of safety belts, not bother with umbrellas on cloudy days, and so on. To conclude, then, contrary to what Domsky says, the selfish bias should have led us to prefer *full* immunity to luck, as guaranteed by intuition 1, over partial immunity to luck, offered by intuition 2.

To make the same point in other words: what we might be considered blameworthy for is wider under intuition 2 than it is under intuition 1. Under intuition 1, we are blameworthy only for what is within our control, while under intuition 2 we are blameworthy both for what is within our control (for example, the careless driving itself) *and* for the unfortunate results of our behavior (for example, the killing of a pedestrian trying to cross the street). Why would anybody choose, *for selfish reasons*, a principle that would make him or her more blameworthy than he or she would be had a different principle been chosen?

I can envisage two possible moves Domsky could make in response to this objection. First, he might argue that although intuition 2 offers us less immunity than intuition 1 in absolute terms, it offers more immunity in relative terms. On intuition 1, factors due to luck neither improve nor discredit our moral status in comparison to others, while given our assumed immunity to luck (the optimistic bias), intuition 2 seems to make us better than others, whom we believe to be far less immune than we are. Maybe this comparative dimension is what Domsky has in mind when he says that intuition 2 "privileges us *at the expense of others* by making us less blameworthy...and by making them more" (471, italics added). The initial plausibility of this response derives from the enormous importance we assign, consciously and subconsciously, to our comparative status. We care very much not simply about how clever we are, but how clever we are in comparison to others, how beautiful we are in comparison to others, and also how blameworthy we are in comparison to others. And since intuition 2 allows us to elevate ourselves "at the expense of others," our self-interest instinct is to prefer it over intuition 1.

---

[13] This is one of the insights of the attribution theory in psychology. Whether events are attributed to luck or to internal factors, that is, to abilities and effort, depends on the type of personality. In particular, depressed people have been shown to ascribe their failures to internal factors, while nondepressives ascribe their failures to external factors (luck and task difficulty). See Nicholas Kuiper, "Depression and Causal Attributions for Success and Failure," *Journal of Personality and Social Psychology*, XXXVI (1978): 236–46.

The trouble with this move is that it makes PSE rely on a psychological assumption that is not substantiated at all by Domsky and, until convincing evidence is presented to show otherwise, I think there is good reason to regard it as false. People do at times prefer a state of affairs that makes their situation worse in absolute terms, only because it improves their relative status in comparison to others, but, thankfully, this does not seem to be the rule. It sounds unlikely that we would take the risk of suffering distressing guilt feelings and severe social condemnation in case *we* unluckily brought about bad outcomes, just to enjoy the satisfaction of looking down at others suffering bad fortune in their turn, which we could then arrogantly interpret as reflecting their low credit.

The second move Domsky could make in response to the above objection is to argue that intuition 2 "stems not just from selfishness but from *misdirected selfishness*" (463). The argument is not entirely clear, but what Domsky seems to be after is the idea that the selfish bias does not necessarily direct the agent to the best results in terms of his or her self interest, hence it can explain our attraction to intuition 2, even if, as I have argued, further reflection shows that we would be better off with intuition 1. But this move takes from the selfish bias all of its explanatory power. This bias was supposed to explain how we hold to some principle (intuition 2) in spite of its being (according to Domsky) blatantly false. PSE was supposed to answer this question by pointing to the psychological mechanisms that distort our cognitive/evaluative judgment, and the selfish bias was one of them. The idea was that because we have a special and uncontrollable attraction to theories and beliefs that favor us, we often fail to notice that these theories and beliefs are false. But if the belief we adopt is one that in fact "favors no one" (463), the entire explanation collapses. Selfish explanations can work only if the preference to be explained can be shown to be the most reasonable path that would be taken by a selfishly rational agent. If the preference, in fact, runs counter to the best interests of the agent, as I tried to demonstrate, there is no good reason to think that it was chosen for selfish reasons.

My conclusion is that if the recommendation to drop intuition 2 is based on PSE, and if this move is supposed to provide us with a solution to the problem of moral luck, then it is unconvincing. However, as explained earlier, it was a mistake in the first place to think that PSE could do the main work in rejecting intuition 2, work that involves discarding the philosophical arguments that led to this intuition. In the next section, I seek to examine whether Domsky has managed to do this.

III. PHILOSOPHICAL ARGUMENTS FOR MORAL LUCK

On Domsky's view, the arguments advanced to solve the problem of moral luck are not simply unsuccessful, but *embarrassingly* so, hence the need for PSE to account for our irrational pull to moral luck. Are these arguments indeed so embarrassing?

Domsky identifies two responses to the problem of moral luck:

> Some philosophers insist that our two intuitions do not actually conflict, and others accept that they conflict but insist that it is a good thing we hold them despite this.[14]

But in the way intuitions 1 and 2 have been formulated by Domsky (445), nobody could possibly argue that they do not conflict; one intuition says that blameworthiness is independent of luck, the other that it is not the case that blameworthiness is independent of luck. How could this fail to be a plain logical contradiction? As an illustration of this approach, Domsky refers to Norvin Richards's paper (*op. cit.*). Yet Richards never denies that intuitions 1 and 2 are contradictory. What he argues is that, in spite of the fact that intuition 2 ought to be rejected, it does capture *something* true about blameworthiness, namely, that our actual judgments about the blameworthiness of people must reflect our epistemic shortcomings, and usually the only way we come to know that some person is morally blameworthy is by what he or she actually did or actually brought about. On my reading of Richards, his *solution* to the paradox of moral luck lies in his unequivocal rejection of intuition 2 as an "ontological" principle (to borrow Domsky's term), while his epistemological addition seeks to restore some rationality to our actual practice of moral criticism, practice that would seem to lose all credibility following the rejection of intuition 2. In other words, the aim of the epistemological argument is not to overcome the contradiction between intuitions 1 and 2, but to soften the apparent contradiction between intuition 1, which Richards fully accepts, and the actual practice of moral criticism. Richards might be unsuccessful in his attempt to save something from intuition 2, but he definitely does not make the mistake of denying the existence of a contradiction between intuitions 1 and 2, nor does he suggest that the partial restoration of intuition 2 through the epistemological argument leads to a complete overlap between judgments of moral criticism justified by this argument and judgments of moral blameworthiness implied by the original, ontological, intuition 2.

---

[14] From "Tossing the Rotten Thing Out" (forthcoming; typescript, p. 1).

Richards's general approach is common to other philosophers too, for example, Hennings Jensen and Judith Thomson (*op. cit.*). Each of them denies intuition 2, thereby denying that there is a real *paradox* in the notion of moral luck. Nevertheless, they try to show, each in his/her own way, why, for practical or epistemological reasons, there is some justification to behave *as if* intuition 2 was true. As a matter of fact, it is not at all clear to me why Domsky should object to this approach, especially as he acknowledges how "extraordinary difficult" it would be to completely drop intuition 2 (464). Given this difficulty, would not it be more realistic to drop only part of it, so to say, that is, the ontological part, while retaining those parts (or those interpretations) of it that *can* be rationally justified, for example, the epistemological?

In any case, as I see no other philosopher who falls under Domsky's heading of thinkers who "insist that the two intuitions do not actually conflict," I must conclude that Domsky is referring here to a straw man or woman.

The second response to the problem of moral luck, says Domsky, consists of "seeing the conflict of intuitions, but being glad for it."[15] While the first response fails to see a logical contradiction where one exists, the latter identifies contradictions but does not care about them, or worse—is *glad* about them. Domsky refers to Margaret Walker as a prominent example of this response (*op. cit.*) but, as far as I can see, she is quite clearly opposed to it. Walker distinguishes *three* positions on moral luck: that it is real but constitutes a paradox; that it is illusory; that it is real but not paradoxical. Her own view is a variant of the third sort of position (*op. cit.*, p. 236), and she expressly rejects the view (which she ascribes to Williams and to Nagel) "that it generates an insoluble problem" (*op. cit.*, p. 238). Hence, I think it is simply false to characterize Walker's position (to which I shall immediately return) as accepting the paradox of moral luck and "being glad" about it.

---

[15] From "Tossing the Rotten Thing Out" (forthcoming; typescript, p. 5).

Another assumed member of this group, on Domsky's view, is Jensen, but, again, I find this classification misleading. Jensen explicitly rejects intuition 2 and makes clear that "actual results beyond agent's control do not determine our verdictive judgments of his moral blameworthiness" (*op. cit.*, p. 139). Thus, Jensen refuses to acknowledge any real paradox in moral luck, and in that respect he belongs—as Walker rightly observes (*op. cit.*, p. 237)—to the same company of Richards, both of whom attempt to show that moral luck is *illusory*.

As a fierce opponent to moral luck, Domsky's real foes are not Jensen, Richards, Thomson, or Michael Zimmerman, all of whom explicitly reject intuition 2, but philosophers like Williams, Nagel, and Walker, who argue for it. Of their arguments, he chooses to deal with one, that developed by Walker, according to which moral luck is a condition for *integrity*. I turn to discuss this argument as described by Domsky, and to show why I think he fails in refuting it.

According to Domsky, Walker believes that "in a world without intuition 2, there would be no blame and therefore no integrity."[16] Domsky objects. In his view, blame can exist without intuition 2, as it can be tied to negligence, rather than to bad outcomes. Therefore, integrity does not depend on intuition 2 either, and we can safely keep the former and get rid of the latter. However, I think that Walker's argument is not that easy to refute. First, she does not argue that in a world without moral luck (a world of "pure agency") there would be *no blame*. There would be blame, but it would be hopelessly limited to those aspects of human behavior which are within human control. A hypothetical world without moral luck would be one with far less blame than a world with moral luck (a world of "impure agency"), not a world without blame altogether.

Second, Walker's argument about the intimate connection between integrity and moral luck relies on her special understanding of integrity, which is rather ignored by Domsky. On Walker's view, what integrity protects is one's moral self, "that center of moral commitments in oneself from which morally fitting and valuable responses flow in a sure and steady way" (*op. cit.*, p. 242). A person of integrity is somebody we can trust and depend upon, especially in trying times where unwanted circumstances impose severe tests. A person of integrity is one who can be trusted to carry his or her commitments through, and to take responsibility over his or her actions, even when he or she has perfectly good excuses to opt out of the burdens imposed thereof, burdens which are so often a result of bad luck. Integrity, argues Walker, is a quality of character "impossible to capture fully without reference to the vicissitudes of moral luck" (*op. cit.*, pp. 241–42). Walker might be wrong in this view, but it cannot be rejected by merely pointing to the fact that blame might be "tied to negligent behavior instead of bad luck," given that on Walker's view integrity is intimately tied to bad luck, not simply to negligence, in the ways just explained.

---

[16] From "Tossing the Rotten Thing Out" (forthcoming; typescript, p. 9).

Domsky's second criticism of Walker concerns her main thesis, namely, that a world without moral luck would be a much worse place to live in. In such a world, argues Walker, a world of pure agency, agents would be responsible only for what is within their control. They would be freed from any responsibility for the unforeseen and the uncontrollable results of their intentional performances. Pure agents enjoy unilateral control of responsibility and are exempt from reparative demands in all areas beyond strict control. That means that they may not be depended on:

> That the very young and old, the weak, the sick, and the otherwise helpless—that is, all of us at some times—depend on the sense of moral responsibility of others unlucky enough to be stuck with the circumstance of their need will not be the pure agents' problem. It is alarming to anticipate life in a world where people routinely and with justification walk away from the harmful, cruel, or even disastrous results that their actions were critical, even if not sufficient, in bringing about (*op. cit.*, p. 247).

To this, Domsky responds by suggesting that the exact opposite is true, namely, that a world with moral luck is much *worse* than a world without. Why so? Because he believes that the extra burdens and responsibilities on unlucky agents in the world of impure agency are bought at the high price of allowing the lucky ones "to fail us completely." People who "routinely and unflinchingly risk other people's lives without a second thought," but are fortunate enough not to cause death or injury, are callously exempt from any responsibility towards the many victims of others' negligence. A world with moral luck is a world in which victims of hit-and-run accidents would be left suffering, because those responsible for their fate, the unlucky hitting drivers, will have continued on their way, while in a world without moral luck, such victims will be assisted by any passing driver. Thus, since a world guided by intuition 2 is a callous one, we have a strong reason to get rid of this intuition.

This move against Walker is surprising and provocative, yet ungrounded. On Domsky's view, anybody who at times behaves negligently ought to help victims of such negligence, even if he himself was fortunate enough not to bring about any bad outcome. Since all of us (or certainly almost all of us) act sometimes in ways that risk other people lives (driving in a less than perfect manner, forgetting the water running in the bath, the fire on the stove, and the like), the upshot is extremely demanding—all of us are under an obligation to provide help to all victims of negligence. Now first of all, this obligation definitely does not follow from intuition 1 (or, for that matter, from the very rejection of intuition 2). Intuition 1 is a limiting principle, clarifying what negligent agents are *not* blameworthy for, that is, the unfortunate outcomes of their (negligent) behavior. It says nothing about the kind of moral burdens that *are* incumbent on agents given their blameworthiness for some wrongful behavior. Since the above obligation to help victims does not presuppose the rejection of intuition 2, it could be adopted by proponents of moral luck too.

This is a good opportunity to clarify that intuition 2 does not replace blameworthiness for negligence by blameworthiness for outcomes, but rather adds the latter to the former. Proponents of moral luck concede that agents are blameworthy for their intentional, wrongful behavior, particularly their unnecessary taking of risks, but they argue that the blameworthiness imposed by such behavior does not exhaust the realm of blameworthiness. We are blameworthy, they contend, not only for voluntary actions and omissions, but for their (unlucky) outcomes *too*. Therefore, *if* blameworthiness for mere negligence involves such demands as those mentioned by Domsky, there is no reason why proponents of moral luck could not acknowledge them too, and just add another package of demands to them, this time deriving from the unlucky outcomes.

Second, the morality of helping victims which is outlined by Domsky seems to smuggle back through the window the very idea of moral luck that was earlier thrown out of the door. On this morality, if I come across victims of a hit-and-run accident in some deserted road, I ought to take responsibility and stop to help them. But why *me*? After all, it is only bad luck that I happened to take this road, at this time, and bump into these victims. If, as assumed by opponents of moral luck, responsibilities must not outrun control, then I should not be seen (by myself or by others) as more responsible for the victims of the accident than anybody else. (On intuition 1, even the hitting driver himself is not more responsible than other negligent drivers for the victims of his negligence, nor should he feel more remorseful than they do.) To say that I do have a special responsibility to help these victims seems to presuppose the role of luck in moral judgment, which was supposed to have been finally denied.

Third, in any case, the morality sketched by Domsky is demanding in a way that makes it unrealistic and alien to human psychology. Domsky expects almost all of us to feel shame and remorse (!) for events to which we have no causal connection, that is, the unwanted results of the negligence of *others*, just because we might have been the ones who caused the harm, or better, just because we ourselves sometimes risk accidentally harming people.[17] And we must all—negligent creatures that we are—take responsibility and provide help to all victims of negligence. While in Walker's world of impure agency, it is the negligent driver who is expected to feel remorse and make amends for the harms he has caused, in Domsky's world, it is all negligent drivers or, indeed (since no causal connection between the negligence and the harm to be repaired is assumed), all negligent agents in whatever sphere who are supposed to feel remorse, and make amends for all regretful results of negligence. The former arrangement strikes me as more practical, more just, and better suited to human nature.

IV. SUMMARY AND CONCLUSION

I argued against Domsky that the problem of moral luck is not captured well when formulated in terms of blameworthiness. In Walker's words, the moral luck issue concerns "a full repertoire of perceptions, judgments, expectations, responses, attitudes, and demands with respect to ourselves and others" (*op. cit.*, p. 238). I presented at some length Walker's argument in favor of a world with moral luck, a world of impure agency, and tried to show the failure of Domsky in refuting her argument. Until this argument (as well as other arguments Domsky does not address in his paper) is refuted, the problem of moral luck cannot be said to have been solved once and forever.

Furthermore, even if Walker (as well as Williams and Nagel) are wrong in extending the realm of responsibility beyond that of controllability, they are definitely not *obviously* wrong, in a way that calls for some kind of psychological account for the assumed breakdown of our rational capabilities. Such a psychological account would be helpful only as an error theory, presuming to explain how people fail to see the obvious. Moreover, the psychological "solution" offered by Domsky fails on its own merits. The selfish bias should have led to the adoption of intuition 1, rather than the adoption of intuition 2.[18]

---

[17] From "Tossing the Rotten Thing Out" (forthcoming; typescript, p. 10).

The solution of the moral luck problem lies in some solid and irresistible philosophical argument for or against moral luck. I do not think Domsky has provided us with such an argument. I am afraid the "stink"—to use again one of his illustrative expressions—instigated by the paradox of moral luck is still in the air. There *is* a door, it is still locked, and we are in need of a key.

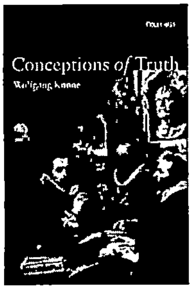
DANIEL STATMAN

Haifa University

---

[18] A different use of psychological studies to shed light on the problem of moral luck has been recently suggested by Edward Royzman and Rahul Kumar, who argue that our "epistemic egocentrism" causes us to incorporate our retrospectively gained knowledge of the outcomes into our understanding of the knowledge that the actor must have had at the time of the action. In their view, this explains why we tend to view the negligent driver who killed a pedestrian as more blameworthy than the driver whose negligence did not lead to such an outcome—see their "Is Consequential Luck Morally Inconsequential? Empirical Psychology and the Reassessment of Moral Luck," *Ratio*, XVII (2004): 329–44.

New from
Oxford



Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

Consciousness
Essays from a Higher-order Perspective
PETER CARRUTHERS

Peter Carruthers's essays on consciousness and related issues have had a substantial impact on the field, and many of his best are now collected here in revised form.

September 2005 paper $29.95 cloth $90.00

Modality and Tense
KIT FINE

Since the 1970s Kit Fine has been one of the leading contributors to work at the intersection of logic and metaphysics. This is his eagerly-awaited first book in the area. It draws together a series of essays, three of them previously unpublished, on possibility, necessity, and tense.

September 2005 paper $29.95 cloth $99.00

Scepticism Comes Alive
BRYAN FRANCES

Bryan Frances presents a highly original argument template for generating new kinds of radical scepticism, ones that hold even if all the clever anti-sceptical fixes defeat the traditional sceptic.

September 2005 $55.00

New in Paperback!
Conceptions of Truth
WOLFGANG KÜNNE

Truth is one of the most debated topics in philosophy; Wolfgang Künne presents a comprehensive critical examination of all major theories, from Aristotle to the present day. He argues that it is possible to give a satisfactory 'modest' account of truth without invoking problematic notions like correspondence, fact, or meaning.

2003 (paper 2005) paper $39.95 cloth $80.00

New in Paperback!
Libertarianism without Inequality
MICHAEL OTSUKA

Michael Otsuka's ingenious and exciting book vindicates left-libertarianism, a political philosophy which combines stringent rights of control over one's own mind, body, and life with egalitarian rights of ownership of the world.

2003 (paper 2005) paper $29.95 cloth $45.00

Burdened Virtues
Virtue Ethics for Liberatory Ethics
LISA TESSMAN

This book is a deeply original and provocative work that engages questions central to feminist theory and practice, from the perspective of Aristotelian ethics. Focused primarily on selves who endure and resist oppression, she addresses the ways in which devastating conditions confronted by these selves both limit and burden their moral goodness, and affect their possibilities of flourishing.

(Studies in Feminist Philosophy)

September 2005 paper $24.95 cloth $65.00

JKC.052705

You can now subscribe to the

THE JOURNAL OF
PHILOSOPHY

using

PayPal

New Rates January 2005

Institutional	$100
Individual	$45
Student/Retired	$20

To learn more about this easy new
option, click on the PayPal icon at
www.jphil.org

Poznań Studies

in the Philosophy of the Sciences and the Humanities

Volume 82

Historical Studies on Abstraction and Idealization

Edited by F. Coniglione, R. Poli & R. Rollinger

Contents:

I. Angelelli, *Adventures of Abstraction*
A. Bäck, *What is Being* qua *Being?*
F. Coniglione, *Between Abstraction and Idealization: Scientific Practice and Philosophical Awareness*
D.P. Henry, *Anselm on Abstracts*
L. Spruit, *Agent Intellect and Phantasms. On the Preliminaries of Peripatetic Abstraction*
R.D. Rollinger, *Hermann Lotze on Abstraction and Platonic Ideas*
R. Poli, *W.E. Johnson's Determinable-Determinate Opposition and His Theory of Abstraction*
M. van der Schaar, *The Red of a Rose. On the Significance of Stout's Category of Abstract Particulars*
C. Ortiz Hill, *Abstraction and Idealization in Edmund Husserl and Georg Cantor prior to 1895*
G.E. Rosado Haddock, *Idealization in Mathematics: Husserl and Beyond*
A. Klawiter, *Why Did Husserl Not Become the Galileo of the Science of Consciousness?*
G. Camardi, *Ideal Types and Scientific Theories*

Managing Editor: dr. Katarzyna Paprzycka · Department of Philosophy · SWPS
ul. Chodakowska 19/31 · 03-815 Warszawa · Poland · Fax: ++48 22 517-9625
E-mail: PoznanStudies@swps.edu.pl · http://PoznanStudies.swps.edu.pl

Publisher: Editions RODOPI, B.V. *USA/Canada:* One Rockefeller Plaza, Suite 1420
New York, NY 10020 · Tel.: 212-265-6460, toll-free (U.S. only) 1-800-225-3993
Fax: 212-265-6402. *All Other Countries:* Tijnmuiden 7 1046 AK Amsterdam
The Netherlands · Tel: ++31 (0) 20 611-4821 · Fax ++31 (0) 20 447-2979
E-mail: orders-queries@rodopi.nl · Web-site: http://www.rodopi.nl

EDITIONS RODOPI B.V. AMSTERDAM – NEW YORK

the review of

metaphysics

a philosophical quarterly

SEPTEMBER 2005 | VOL. LIX, No. 1 | ISSUE No. 233 | $15.00

articles

ALLEGRA DE LAURENTIIS	*Metaphysical Foundatio of the History of Philosop*
JAMES D. REID	*Ethical Criticism in Heidegger Early Freiburg Lectur*
FREDERICK J. CROSSON	*Esoteric versus Latent Teachi*
JOHN GARDNER	*Wrongs and Faul*
MARK C. MURPHY	*The Common Goo*
ELIZABETH C. SHAW AND STAFF	*Summaries and Commen*

book reviews *philosophical abstracts* *announcemen*

Individual Subscriptions $35.00 Institutional Subscriptions $55.00 Student/Retired Subscriptions $20

VISA AND MASTERCARD ACCEPTED

Jude P. Dougherty, Editor
The Catholic University of America, Washington D.C. 20064
email: mail@reviewofmetaphysics.org

ALP The Amherst Lecture in Philosophy

Amherst College congratulates

J. David Velleman

Professor of Philosophy, New York University

who will deliver the inaugural

Amherst Lecture in Philosophy

Thursday, March 9, 2006

For more information
on this new annual lecture
and digital publication go to

www.amherstlecture.org

CAMBRIDGE

OUTSTANDING SCHOLARSHIP FROM CAMBRIDGE

Now in paperback...

The Cambridge History of Hellenistic Philosophy

Edited by Keimpe Algra, Jonathan Barnes, Jaap Mansfield, and Malcolm Schofield

The Cambridge History of Political Thought

$48 00*· Paperback. 0-521-61670-0 c 916pp

Aquinas: Disputed Questions on the Virtues

Edited by E.M. Atkins and Thomas Williams

Cambridge Texts in the History of Philosophy

$75.00· Hardback. 0-521-77225-7: 344pp
$26 99. Paperback. 0-521-77661-9

The Political Philosophy of Cosmopolitanism

Edited by Gillian Brock and Harry Brighouse

$70.00· Hardback: 0-521-84660-9· 270pp
$24.99· Paperback: 0-521-60909-7

Cognition and the Brain

The Philosophy and Neuroscience Movement

Edited by Andrew Brook and Kathleen Akins

$90.00: Hardback: 0-521-83642-5: 452pp

Plato's Cosmology and its Ethical Dimensions

Gabriela R. Carone

$70 00· Hardback. 0-521-84560-2: 296pp

The Cambridge Companion to Foucault

Second Edition

Edited by Gary Gutting

Cambridge Companions to Philosophy

$70.00· Hardback: 0-521-84082-1· 488pp
$28.99· Paperback· 0-521-60053-7

Wittgenstein's 'Tractatus'

An Introduction

Alfred Nordmann

Cambridge Introductions to Key Philosophical Texts

$70.00: Hardback: 0-521-85086-X: c.216pp
$24.99· Paperback: 0-521-61638-7

Understanding Ethical Failures in Leadership

Terry L. Price

Cambridge Studies in Philosophy and Public Policy

$65.00· Hardback· 0-521-83724-3. 240pp
$24.99. Paperback. 0-521-54597-8

Now in paperback...

The Cambridge History of Greek and Roman Political Thought

Edited by Christopher Rowe and Malcolm Schofield

The Cambridge History of Political Thought

$48.00*· Paperback: 0-521-61669-7· c.745pp

Biodiversity and Environmental Philosophy

An Introduction

Sahotra Sarkar

Cambridge Studies in Philosophy and Biology

$75.00: Hardback· 0-521-85132-7: c.220pp

Heidegger on Ontotheology

Technology and the Politics of Education

Iain D. Thomson

$70 00: Hardback: 0-521-85115-7: 224pp
$24 99: Paperback: 0-521-61659-X

The Cambridge Companion to Berkeley

Edited by Kenneth P. Winkler

Cambridge Companions to Philosophy

$70.00*: Hardback: 0-521-45033-0. c.472pp
$28.99*: Paperback. 0-521-45657-6

Symmetry and Its Discontents

Essays on the History of Inductive Probability

S.L. Zabell

Cambridge Studies in Probability, Induction and Decision Theory

$70.00: Hardback. 0-521-44470-5. 288pp
$24.99: Paperback. 0-521-44912-X

**Prices subject to change.*

www.cambridge.org/us/philosophy/

CAMBRIDGE
UNIVERSITY PRESS

Datta
1·3·06

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NUMBER 9
SEPTEMBER 2005

page

437 *The Wrong Kind of Reason* — Pamela Hieronymi
458 *Deception Unraveled* — Alan Strudler

COMMENTS AND CRITICISM

474 *New Distinctions, Same Troubles: A Reply to Haji and McKenna* — David Palmer

BOOK REVIEWS

483 Tim Maudlin: *Truth and Paradox: Solving the Riddles* — Graham Priest
487 NEW BOOKS

***Published by* The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Christopher Peacocke, Philip Pettit, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721. You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2005

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.
POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2005 by the Journal of Philosophy, Inc. ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY
VOLUME CII, NO. 9, SEPTEMBER 2005

THE WRONG KIND OF REASON*

A good number of people currently thinking and writing about reasons identify a reason as *a consideration that counts in favor of an action or attitude.*[1] I will argue that using this as our fundamental account of what a reason is generates a fairly deep and recalcitrant ambiguity; this account fails to distinguish between two quite different sets of considerations that count in favor of certain attitudes, only one of which is the "proper" or "appropriate" kind of reason for them. This ambiguity has been the topic of recent discussion, under the head "the wrong kind of reasons problem."[2] I

---

* Earlier versions of this material benefited from conversation or commentary from Tyler Burge, Sarah Buss, John Carriero, Sonny Elizondo, Mark Greenberg, Aaron James, Sean Kelsey, Niko Kolodny, Clayton Littlejohn, Matthew McAdam, T.M. Scanlon, and audiences at Princeton University, the University of Michigan, and the Pacific Division Meeting of the American Philosophical Association.

[1] See, for example, Jonathan Dancy, *Practical Reality* (New York: Oxford, 2000); Derek Parfit, "Rationality and Reasons," in Dan Egonsson et al., eds., *Exploring Practical Philosophy* (Burlington, VT: Ashgate, 2001), pp. 17–39; Joseph Raz, *Practical Reason and Norms* (London: Hutchinson, 1975; reprint, Princeton: University Press, 1990); Raz, *Engaging Reason: On the Theory of Value and Action* (New York: Oxford, 1999); T.M. Scanlon, *What We Owe to Each Other* (Cambridge: Harvard, 1998); J. David Velleman, *The Possibility of Practical Reason* (New York: Oxford, 2000). Obviously, these writers have in mind what are sometimes called "justifying" reasons. What are sometimes called "explanatory" reasons would require different treatment. John Broome takes the explanatory role of reasons as primary, and identifies a justificatory (or "normative") reason as a fact that explains an ought claim. See Broome, "Reasons," in R. Jay Wallace et al., eds., *Reason and Value: Themes from the Moral Philosophy of Joseph Raz* (New York: Oxford, 2004), pp. 28–55.

[2] See, for example, Jonas Olson, "Buck-Passing and the Wrong Kind of Reasons," *The Philosophical Quarterly*, LIV, 215 (2004): 295–300; Justin D'Arms and Daniel Jacobson, "Sentiment and Value," *Ethics*, CX (2000): 722–48; Wlodek Rabinowicz and Toni Rønnow-Rasmussen, "The Strike of the Demon: On Fitting Pro-Attitudes and Value," *Ethics*, CXIV (2004): 391–423. The "wrong kind of reasons" problem, as identified by these authors, concerns reasons for "valuing" or for those "pro-attitudes" or "sentiments" that are thought to be appropriate when their object is valuable. So understood, the wrong kind of reasons problem is thought to challenge the viability of a

0022-362X/05/0209/437–57

will suggest that confusion about "the wrong kind of reason" will be dispelled by changing our account of what a reason is. While I agree both that reasons are considerations and that certain considerations count in favor of actions and attitudes and thus provide reasons for them, we should not understand "counting in favor of an action or attitude" as the fundamental relation in which a consideration becomes a reason. Rather, we would do better to think of a reason as *a consideration that bears on a question.* This alternative will allow us to distinguish "kinds" of reasons by distinguishing between kinds of questions on which a consideration can bear, and it will allow us to distinguish the "right" kind of reasons for certain attitudes from the "wrong" kind, by allowing us to consider the relation between the question on which the consideration bears and the attitudes of which it counts in favor.

REASONS AS CONSIDERATIONS IN RELATION

Before explaining how the original account generates the ambiguity (and how the alternative avoids it), I will orient discussion by noting some points of commonality between the original formulation and my alternative. Both accounts understand reasons as considerations standing in some sort of relation. In agreeing that reasons are considerations, it may seem I am taking sides in a debate about reasons for action. There has been considerable debate about whether reasons for action are mental states—beliefs and desires, in particular—or rather are "considerations," that is, the propositions, facts, states of affairs, events, or objects that might serve as the *content* of such mental states, that which the beliefs or desires are about.[3] Which side one

---

certain account of value (the "fitting-attitudes" or "neo-sentimentalist" account). In fact, I believe this discussion involves several distinguishable issues, only one of which I will address: I will consider the question of how to distinguish the "right" reasons for an attitude from the "wrong" kind. I will not focus particularly on "valuing" or on "pro-attitudes," because, as noted by D'Arms and Jacobson, as well as by Parfit and Christian Piller (references below), there is a quite general problem about identifying the appropriate reasons for attitudes, a problem that is not restricted to reasons for those attitudes involved in valuing. I hope to resolve this *broader* problem about the "wrong kind of reason." Even if I am successful, problems about value will remain. I consider these briefly late in the paper.

[3] Donald Davidson is often associated with the view that reasons for actions are belief/desire pairs. See, for example, Davidson, "Actions, Reasons, and Causes," in *Essays on Actions and Events* (New York: Oxford, 1980), pp. 3–19. It is not at all clear to me, however, that Davidson had the distinction in question in mind. The view the reasons for action are belief/desire pairs has been recently defended by, among others, Michael Smith. See, for example, "The Possibility of Philosophy of Action," in Jan Bransen and Stefaan E. Cuypers, eds., *Human Action, Deliberation, and Causation* (Boston: Kluwer, 1998), pp. 17–41, and *The Moral Problem* (Cambridge: Blackwell, 1994). The contrasting view has been argued by, among others, Stephen

favors in this debate typically turns on whether one is trying to explain action or rather trying to determine how or whether to act. While it can seem obvious that actions are to be explained by facts about the mental life of the agent, it can seem equally obvious that reasons which settle the question of how or whether to act are facts of a more general sort—including facts about the weather or the time or how much something costs or how far away the movie theater is. For present purposes, I hope it can be agreed on all sides that, once we widen the field to include not only reasons for action but also reasons for beliefs and desires (and even reasons for weather patterns), reasons cannot be restricted to mental states. But "considerations," broadly construed—facts about the time or the location of the movie theater, or facts about the perceptual system of a believer, the desires of an agent, or the evolution of the species—could, it seems, be reasons of a more general sort. So I will presume that reasons can be understood as considerations, broadly construed.[4]

I also share with the original account the assumption that considerations do not provide reasons in isolation, so to speak. The fact that my eyesight is poor or that my office is on the third floor is not yet a reason. It is just a fact. The same fact could provide many different reasons. The fact that my office is on the third floor might be a reason to take the elevator, the reason it is warm in here, the reason my students do not come to see me, a reason to think I am highly favored, a reason to feel guilty, and so forth. Considerations become reasons only by standing in some sort of relation.[5]

Darwall, Thomas Nagel, Raz, Warren Quinn, and Scanlon. A recent, extended, interesting discussion can be found in Dancy, *Practical Reality.* I consider reasons for action in "Reasons for Action" (in progress).

[4] I am at this point overlooking disagreement about whether a reason must be a fact, whether it can be a purported fact, whether it is a proposition, or whether it is the kind of thing that can be the subject matter of a proposition, and so on. I am taking "consideration" loosely enough to avoid commitments on these issues. (From the point of view of the speaker, a reason is always a fact, a true proposition or its subject matter. However, reasoners are often mistaken. When a reasoner is mistaken about a matter of fact, she could mistakenly take a consideration that is not [in fact] a fact to be a reason. In such a case, the reasoner acts on reasons that are not facts. Some will say they are, therefore, not "really" reasons, or not "good" reasons. While I would insist that such nonfacts really are *that person's* reasons, it seems right to say that they are not good reasons, and, further, that to take something as one's reason is to take it to be, in this sense, a good reason—that is, that reasons must at least be taken to be facts, from the point of view of the one for whom they are reasons.)

[5] Raz notes that it is natural to think of a reason for action as a relation between "a fact which is the reason...a (class of) person(s) and an action(-type)" ("Agency, Reason, and the Good," in *Engaging Reason,* p. 22). See also Raz, *Practical Reasons and Norms.* John Skorupski gives a similar, though more complex, account of what he calls "the reasons relation." He analyzes reasons for action as follows: "the fact that p gives [person] x reason of degree d at time t to ψ, where ψ ranges over action-types." He

My alternative differs from the original formulation in the relation in which it takes a reason to stand.

COUNTING IN FAVOR OF ATTITUDES

Taking a reason to be a consideration that counts in favor of an action or attitude generates a fairly deep and recalcitrant ambiguity in an important range of cases—it generates the wrong kind of reasons problem.[6]

The problem appears most readily in reasons for believing. A consideration can count in favor of believing in two quite different ways: it can show the content of the belief true or it can show the belief, as an attitude, in some way good (useful, convenient, and so on) to have. The fact that the butler wanted revenge is a reason, of the first sort, for believing the butler is guilty; the fact that believing him guilty is the only way to save your life is a reason, of the second sort, for believing he is.[7]

The ambiguity also appears in reasons for intending. A consideration can count in favor of intending to ϕ either by bearing on whether to ϕ or by showing an intention to ϕ is in some way good to have. Consider the politics of deterrence. You might find yourself convinced that an intention to retaliate is good to have—for its deterrent ef-

---

provides a parallel account of reasons for belief, "the fact that *p* gives [person] *x* reason of degree *d* at time *t* to believe that *q*," and for what he calls "reasons to feel," "the fact that *p* gives [person] *x* reason of degree *d* at time *t* to feel ϕ, where ϕ ranges over types of feelings (that is, emotions, moods, and desires)" ("Reasons and Reason," in Garrett Cullity and Berys Gaut, eds., *Ethics and Practical Reason* (New York: Oxford, 1997), pp. 345–67, on p. 346). According to Scanlon, the relevant relation is "counting in favor of" some other item. Scanlon takes a reason to be a consideration that counts in favor of something (*What We Owe to Each Other*, pp. 17–18). As Scanlon's discussion proceeds, one learns that the particular item reasons seem to count in favor of are what he calls "judgment-sensitive attitudes," attitudes such as belief, desire, intention, fear, resentment, or jealousy. So long as any action involves an intention, Scanlon's account will cover both reasons for actions. Parfit puts his similar view quite simply: "Facts give us reasons...when they count in favour of our having some belief or desire, or acting in some way" ("Rationality and Reasons," p. 18). Dancy says, "a reason is a consideration that speaks in favor of action (or belief)" (*Practical Reality*, p. 144). However, Dancy provides some further reflection on what he calls the "ground" for a reason. Whereas, for Scanlon, the notion of "counting in favor of" is primitive (and values are understood through this notion, via the account of reasons), for Dancy reasons are grounded in values, and thereby count in favor of things. My account is closer to Scanlon's, in that it does not make explicit use of a conception of value.

[6] It also creates difficulty for the already difficult project of explaining action. I consider the explanation of action, and argue for my preferred account of a reason, in "Reasons for Action."

[7] Why would believing the butler did it save your life? Perhaps the mob has made credible threat to kill you if you do not give convincing testimony of his guilt, and perhaps you are a terrible liar.

fects—without thereby thinking retaliating is to be done.[8] Reasons for wanting are similarly ambiguous: the fact that you must win tonight's game in order to advance to the championships is a reason to want to win; however, the fact that lacking a desire to win would disappoint or offend your teammates is also a reason—of a different sort—for wanting to win.

We could continue to draw out an ambiguity in reasons for fearing, resenting, envying, admiring, forgiving, and trusting, among others.

Both Derek Parfit and Christian Piller have noticed the breadth of this ambiguity. Parfit labels it as a distinction between "object-given" and "state-given" reasons, while Piller uses the labels "content-related" and "attitude-related."[9] Neither gives a precise account of how to draw the distinction. Both simply suggest that object-given or content-related reasons are provided by or have something to do with the object or content of the attitude, while state-given or attitude-related reasons are provided by or have to do with the state or attitude itself.[10]

As noted by Piller, some might attempt to avoid the problem by denying that attitude-related reasons really are reasons for the attitude. Rather, the supposed attitude-related reasons for believing *p*, say, are really reasons for *wanting* or *trying* to believe *p*. The only reasons that really count in favor of believing *p* are those that show *p* true.[11]

While it is obviously true that only certain reasons are the "real" or "appropriate" kind of reasons for believing—that others are of the wrong kind—and while I happily agree that attitude-related reasons for one attitude can serve as content-related reasons for another attitude (namely, a second-order attitude whose content concerns the first), the problem lies in saying why the attitude-related reasons are not "really" reasons for the attitude in question. Reasons, on the original formulation, are considerations that count in favor of an action or attitude. Considerations that show a belief good to have surely "count in favor of" believing, quite independently of whether they show the belief true. Considerations that show something good about believing *p* count in favor of believing *p* in just the same way that considerations that show something good about attending lecture count in favor of attending lecture. Likewise, the fact that supposing *p* would advance the argument counts in favor of supposing *p* and is, thereby, a reason for supposing *p*. But if the fact that supposing *p* would advance the argument can be a reason to suppose *p*, the fact that believing *p* would advance my career should, for all that has been said, be a reason for believing *p*. If a reason is simply a consideration that counts in favor of an action or attitude, attitude-related reasons should qualify as reasons for the attitude of which they count in favor.[12]

---

[8] Cf. Gregory Kavka, "The Toxin Puzzle," *Analysis*, XLIII (1983): 33–36.

[9] See Parfit, "Rationality and Reasons"; Piller, "Normative Practical Reasoning," *Proceedings of the Aristotelian Society*, Suppl. Vol. LXXV, 1 (2001): 195–216, and "Content-Related and Attitude-Related Reasons for Preferences," *Philosophy* (forthcoming).

[10] Parfit introduces the distinction when discussing reasons for desires, and says, "Of our reasons to have some desire, some are provided by facts about this desire's object. These reasons we can call object-given.... Other reasons to want something are provided by facts, not about what we want, but about having this desire. These reasons we can call state-given" ("Rationality and Reasons," p. 21–22). Piller, who notes that he does not aim to give a precise definition, says, "Take some propositional attitude for which there could be reasons, like wanting *p*, intending that *p*, or believing that *p*. Something is a content-related reason if it refers to some appropriate feature of *p*. For example, believing that *p* is useful would be a content-related reason for wanting that *p*. In other words, *p*'s usefulness is a consideration that speaks in favor of wanting that *p*. Believing that *p is the best explanation of q*, where *q* is something believed to be true, is a content-related reason for believing that *p*.... Something is an attitude-related reason for a propositional attitude if it refers to some appropriate feature of the attitudes as a whole and not only to a feature of the content of the attitude. Take wanting that *p*. One of the features of the attitude as a whole would be that *wanting that p is useful*. Believing that *wanting that p is useful* would then be an attitude-related reason for wanting that *p*" ("Normative Practical Reasoning," pp. 204–05). The account lacks precision because of its reliance on 'appropriate'.

[11] See Piller, "Normative Practical Reasoning," p. 206. He argues against this view in "Content-Related and Attitude-Related Reasons for Preferences."

Further, for certain propositional attitudes, such as supposing, imagining, and remembering, the "real" reasons seem to be attitude-related, while reasons that bear on the content are reasons for a different attitude. Suppose, for example, that you are asked to imagine there's no heaven. We can draw a distinction between considerations that bear on the content of the imagining and those that instead simply show imagining in some way good to do. Divine revelation or the human capacity for wishful thinking would bear on the content of the imagining—whether there is a heaven—while the fact that you have been asked to imagine it, or the fact that it's easy if you try, bear on whether to imagine. Here, the "real" reasons for the attitude seem to be the attitude-related reasons, while reasons that bear on the content seem to be the proper reasons for a different attitude—reasons to believe.

Some might try to show that the content-related reasons are the only "real" reasons for believing by appealing to standards of justification or rationality: although the attitude-related reasons *count in favor* of believing, they are not really reasons for believing, because they do not *justify* or *rationalize* belief. This appeal, simply so put, is unsatisfactory. Once we grant that reasons are considerations that count in favor of an action or attitude, attitude-related reasons will themselves provide corresponding standards of rationality and justification.[13] If you believe that the butler did it because believing this will save your life, then there is a sense in which you have good reason to believe it, and so a sense in which you are justified and even rational in doing so. Of course, this is justification of the wrong kind. But we need some account of why it is the wrong kind. Thus, simple appeal to notions of justification or rationality seems to rely on, rather than elucidate, the distinction we are after. For an appeal to justification or rationality to be satisfactory, we would need an independent account of why only *some* of the reasons that count in favor of an attitude justify or rationalize it—which is just to say, we would need an answer to the wrong kind of reasons problem.

---

[12] Rabinowicz and Rønnow-Rasmussen agree with me in this: "Surely, if we are supposed to have reasons for *actions* when the actions have useful effects or are valuable for their own sake, why shouldn't we have reasons for *attitudes* in comparable circumstances?" ("Strike of the Demon," p. 413). See also Piller, "Normative Practical Reasoning," pp. 211–15.

So we are left with the ambiguity. As long as a reason is simply a consideration that counts in favor of an attitude, we are left without an obvious way either to draw a useful distinction between these very different sorts of reasons or to say why one of them seems to be the "real" sort of reasons.

AN ALTERNATIVE ACCOUNT

In light of this problem, I suggest we abandon the thought that a reason is a consideration that counts in favor of an action or attitude and look for a different account.[14]

To start reflection, we can note that, most generally, a reason is simply an item in a piece of reasoning. What is a piece of reasoning?

---

[13] Gibbard notes the connection between reasons and rationality in *Wise Choices, Apt Feelings* (Cambridge: Harvard, 1990). He distinguishes the rationality of "*feeling* or *believing* something" from the rationality of "*wanting* to feel or believe it" at p. 37. D'Arms and Jacobson discuss his treatment. See "Sentiment and Value," pp. 412–15. Parfit considers the matter of rationality in "Rationality and Reasons" at p. 24. Both treatments are discussed by Rabinowicz and Rønnow-Rasmussen, "Strike of the Demon," pp. 412–15.

For an argument that pragmatic considerations do not rationalize belief, see Thomas P. Kelly, "The Rationality of Belief and Some Other Propositional Attitudes," *Philosophical Studies*, CX (2002): 163–96. For an argument that epistemic rationality does not reduce to instrumental rationality, see Kelly, "Epistemic Rationality as Instrumental Rationality: A Critique," *Philosophy and Phenomenological Research*, LXVI, 3 (2003): 612–40.

[14] As noted, Broome offers a different account, according to which reasons explain, and "normative" reasons explain ought facts. He then argues against the view that counting in favor is the basic normative notion saying, "it cannot be, because it is complex. It incorporates the two elements of normativity and explanation" ("Reasons," p. 41). It seems to me that this argument will convince only if one is already persuaded that a reason is a fact that explains an ought claim. I hope to have provided reasons to be unsatisfied with the notion of "counting in favor of an action or attitude"

Remaining very vague, we can at least contrast a piece of reasoning with a mere collection of claims, propositions, or concepts, or a mere series of thoughts, contents, or representations, by noting that a piece of reasoning has some direction to it—it is the sort of thing that might have a conclusion, in something like the way an argument has a conclusion.[15] And the items in a piece of reasoning (again unlike the items in a mere collection of claims or propositions or in a mere series of thoughts or contents) stand in certain relations to each other—they bear what we might call "rational relations" to one another—and thereby also bear on the conclusion. Further, reasoning seems called for just in case a conclusion is unknown or called into question. So I suggest, for consideration, the following account of a reason: a reason is a consideration that bears on a question.[16]

This account differs from the original formulation in taking the fundamental relation in which a consideration becomes a reason to be a relation to a question, rather than to an action or attitude. The new formulation will allow us to distinguish between the content- and attitude-related reasons by distinguishing between questions. Perhaps surprisingly, distinguishing content- and attitude-related reasons will not allow us to identify the "right kind" of reasons for an attitude. However, on the new formulation, the relation between attitudes and reasons is mediated by a question. Thus, this formulation requires us to consider the relation between the question on which the consideration bears and the attitude of which it counts in favor. Considering this relation will help to identify the right kind of reason.

SORTING REASONS: CONTENT- AND ATTITUDE-RELATED[17]

A consideration can "count in favor" of certain attitudes by bearing on either of two distinct questions. Consider, again, belief. A consideration can count in favor of believing *p* by bearing on whether *p* or by bearing on whether the belief that *p* is in some way good to have. We can draw the distinction between content- and attitude-related reasons for belief accordingly. Reasons that bear on the first question

---

independently of such an account.

[15] Gilbert Harman has done much to differentiate reasoning from argument. See, for example, *Change in View* (Cambridge: MIT, 1989). I do not mean, here, to assimilate them, nor to choose between an account of reasoning as something concrete agents undertake in time and space and an account of reasoning as a set of abstract relations.

[16] One could say, "a consideration that bears on a conclusion." I do not think there would be any relevant difference, though I find the idea of answering a question more intuitive for capturing the activities of rational agents.

[17] I also discuss how to draw this distinction in "Controlling Attitudes," *Pacific Philosophical Quarterly* (forthcoming).

(whether *p*) are content-related reasons for a belief that *p*; reasons that bear on the second question (whether the belief that *p* is in some way good to have) are attitude-related reasons for that belief. Likewise for intention: a consideration can count in favor of intending to ϕ by bearing on whether to ϕ or by bearing on whether the intention to ϕ is good to have. Content-related reasons for intending to ϕ bear on the first question (whether to ϕ); attitude-related reasons bear on second (whether the intention to ϕ is good to have).

We can generalize the method: attitude-related reasons count in favor of an attitude by bearing on the question of whether the attitude is, in some way, good to have. Content-related reasons somehow manage to count in favor of an attitude by bearing on a different question—a question that typically does not mention the attitude itself, but concerns its content.

Notice two features of this distinction, so drawn: first, we have not given a general way to specify the content-related question. In the central cases of belief and intention, the question is clear enough (whether *p* or whether to ϕ, respectively). However, when generalized to other attitudes (desiring, resenting, admiring, trusting), the question is not obvious, and we have not given a way to identify it. Rather, we have specified the attitude-related question in a general way (whether the attitude is in some way good to have); we then identified the content-related reasons as those that somehow manage to count in favor of the attitude by bearing on some *other* question, presumably one which has to do with the attitude's content.[18]

Second, note that, even in those cases in which the content-related question is clear enough (the cases of belief and intention), the distinction between content- and attitude-related is not exclusive: the answer to one of these questions can bear on the other question, and thus reasons that provide an answer to one question can, in turn, become reasons of the other sort. For example, quite often (though not always), it is good to have true beliefs. Thus, by showing the content of a belief to be true, content-related reasons might also thereby show the belief is good to have. For beliefs with certain contents (such as the admittedly odd, "This very belief is good to have") reasons showing the belief good to have may also, thereby, bear on whether its content is true.[19] Likewise, it is almost always good to intend to do what there is reason to do; so content-related reasons for intentions are almost always also attitude-related reasons for that intention (an exception would be the problematic cases in which there is reason to do something only if you do it unintentionally). Finally, reasons that show an intention good to have can readily bear on whether to act. Your complaint that I have no intention of attending your party can be a reason for me to attend.[20] In general, the considerations that answer the content-related question might also, thereby, show the attitude good to have, while considerations that show the attitude good to have might (at least in principle) bear on the content-related question.

---

[18] This formulation will surely capture a broader range of reasons than the "right kind." For example, a consideration might count in favor of resenting by bearing positively on whether the person resented would be offended by being (instead) forgiven. This reason counts in favor of resenting by bearing on a question other than whether resenting is good to do (though it will, eventually, come to bear on that question), which question has to do with the content of the attitude (the person). One might try to say that the "right kind" of reasons are those that somehow manage to count in favor of an attitude *without* bearing on the question of whether the attitude is in some way good to have. But this will be too restrictive, as shown by the next paragraph.

Previous attempts to resolve the wrong kind of reasons problem have relied on something like the content-/attitude-related distinction. But, since this distinction is not exclusive, it will not do the job; it cannot distinguish between the "right" kind of reasons for believing or intending and those that are somehow inappropriate.[21] Moreover, because we do not have a general way to specify the content-related question, the distinction cannot be made exclusive. We would like to identify those that bear *only* on the attitude-related question as the "wrong kind" of reason, while identifying any consideration that bears on the content-related question (regardless of whether it also bears on the attitude-related question) as the "right kind." But we cannot do this without specifying the content-related question. Thus, to solve the wrong kind of reasons problem, generally, we must be able to specify the question on which the right kind of reason bears. We can do this, I now suggest, by considering the relation between *settling a question* and *forming or revising an attitude.*

---

[19] In addition to the peculiar, "This belief is good to have," consider the following case: Suppose you are interviewing people to run your propaganda campaign, and you find yourself thinking it would be good to believe that a certain person would be best for the job. Moreover, you find yourself thinking the belief would be good to have as a result of the candidate's own activities. In this case, the fact that it would be good to believe that she will be an effective leader of your campaign might, itself, be a reason that shows that she will. Here, *p* is a proposition whose truth you think is made more likely by the fact that you have a reason for wanting to believe it. I consider this case in "Controlling Attitudes."

[20] If one doubts that the reason you have given me, here, is really attitude-related, we can change the case. Suppose you have a reliable intention detector. As we sit in your lab, you are interested in whether intentions for different kinds of action produce differing patterns on your instrument, and so ask me to intend to drink a glass of water. I know what to do: I drink the water. My reason for drinking is that I know that, by deciding to drink, I will produce in myself the desired intention. Kavka's "Toxin Puzzle" provides an interesting case in which reasons that show the intention good to have do not, I would argue, bear on whether to act.

[21] Rabinowicz and Rønnow-Rasmussen initially attempt to solve the wrong kind of reasons problem by appeal to the fact that certain properties both appear in the "content" of the pro-attitude and, at the same time, justify the attitude ("Strike of the Demon," p. 414). They are relying on the thought that that the right kind of reasons for valuing must appeal to some property of the thing valued. But this runs into counterexamples, because not all properties of a thing are (really) reasons to value it. They imagine an evil demon who requires you to admire him "on account of his determination to punish [you]" if you do not. He seems to have given you a reason to admire him that is of the wrong kind. Yet, he has appealed to a property of the object (himself) as the basis for the admiration.

SORTING REASONS: CONSTITUTIVE AND EXTRINSIC[22]

Note that, for certain attitudes, settling a question amounts to forming the attitude. Suppose you take certain considerations to bear positively on whether *p*. In fact, you take those considerations to settle the question of whether *p*. You are convinced by those reasons. If you find the reasons that (you take to) bear positively on whether *p* to be convincing, and so settle for yourself the question of whether *p*, you have thus, *ipso facto,* formed a belief that *p*. I call the reasons that (are taken to) bear on whether *p* the "constitutive" reasons for believing *p*. Finding them convincing and so settling the question on which they bear amounts to believing.[23]

Using Broome's understanding of what a reason is (that which explains an ought fact), Olsen suggests that the right kind of reasons are those which explain why you ought to have an attitude *without* appealing to "properties of the attitude in question" ("Buck-Passing and the Wrong Kind of Reasons," p. 299). But surely there could be cases in which a fact that (properly) explains why one ought to adopt an attitude makes reference to the very attitude. Perhaps one ought only to love (romantically) those who respond well to one's love. Whatever one thinks of this as an account of love, it should not be ruled out by a general formula about what could qualify as the "right kind of reasons" for an attitude.

Both these attempts, then, fail for the same reason. They attempt to specify the "wrong" kind of reason by appeal to whether the reason appeals to facts about the attitude or to facts about its object. But, as I will suggest in the text, finding the "right" kind of reason requires attending instead to agency, that is, asking, not about what the reason appeals to, but about what taking it to be sufficient amounts to.

[22] I introduce and discuss this distinction at some length in "Controlling Attitudes."

[23] The reasons are not literally constitutive of the belief. You could maintain the belief while forgetting the reasons. Rather, they are the reasons that bear on the question, the settling of which is constitutive of believing.

Note also, the constitutive reasons are those that *bear or are taken to bear* on the question, the settling of which amounts to believing. There is a question, the settling of which amounts to believing *p* (the question of whether *p*). People can disagree about which considerations really do bear on that question, and to what extent. The distinction between the "right" and "wrong kind" of reasons for believing should not, in my view, settle that disagreement. Rather, once we identify the right question, Consider, now, the *remaining* reasons for believing *p*—those which (are taken to) count in favor of believing *p* independently of whether *p*. Finding these reasons convincing does not amount to believing *p*. Suppose that your life depends on your ability to give convincing testimony that the butler did it, but you do not believe he did, and you are a terrible liar. Now, you might think that you have overwhelming reason showing it good to believe the butler did it—it would save your life. Yet, the fact that you take these reasons to be fully convincing does not itself show that you believe the butler did it—because it does not show that you have settled the question of whether he did. Rather, by finding these reasons convincing you have settled the question on which *these* reasons bear, namely, the question of whether it would be good to believe the butler did it. Thus, by finding these reasons convincing, you form a *second-order* belief *about* the belief that the butler did it: you believe it would be good to believe he did. (You might also desire to believe he did, or, if you have the means, you might intend to bring it about that you believe the butler did it.) Since finding these reasons convincing implies nothing about whether you have the belief of which they count in favor, I call them "extrinsic" reasons for that belief. These extrinsic reasons for believing *p* are constitutive reasons for a quite different attitude—a second-order attitude about the belief that *p*.[24]

This distinction between constitutive and extrinsic reasons for a belief that *p* marks the distinction between the "right kind" and the "wrong kind" of reasons for believing *p*. The right kind of reasons for believing *p* are those that (are taken to) bear on whether *p*—that is, those that (are taken to) bear on the question, the settling of which amounts to believing. Extrinsic reasons are not "really" reasons for believing *p*, we can say, because they are not the kind of reasons which, simply by finding convincing, one would believe *p*. Rather, by

---

we have identified the right kind of reason. There will be further, substantive issues about which among these reasons are good ones. I return to this point in the body below.

[24] The constitutive/extrinsic reason distinction is closely related to the content-/attitude-related distinction. I suggest that we can identify the content-related question(s) as that (those) the settling of which amounts to forming the attitude. From within a single point of view, the class of constitutive reasons for believing will be identical to the class of content-related reasons for believing—and will include those attitude-related reasons which, by showing the belief is good to have, also thereby bear on the truth of its content. Extrinsic reasons are simply those that are not constitutive. As it turns out, the extrinsic reasons will be identical to the *remaining* attitude-related reasons: they count in favor of the belief by bearing on whether the belief is in some way good to have. (I see no other question on which a reason could bear and so count in favor of the belief.)

finding them convincing, one will form a different attitude—a second-order attitude about the belief that *p*. Extrinsic reasons for a belief that *p* are constitutive reasons for a belief about the belief that *p*.

I suggest that we can, in general, distinguish the right kind of reason for an attitude from the wrong kind, not by appeal to whether the reason is content- or attitude-related—not, that is, by appeal to whether the reasons show the attitude good to have, or whether appreciating the reason requires some sort of reflection upon the attitude itself, or whether it appeals to a feature of the content of the attitude—but rather by considering how we form the attitude, when we do so for reasons—that is, by considering our agency over that attitude. The "right kind of reasons" (are taken to) bear on a question, the settling of which amounts to forming the attitude. The wrong kind of reasons do not bear, or are not taken to bear, on that question.

It is worth noting that, while drawing the constitutive/extrinsic distinction will solve the "wrong kind of reasons" problem, it will not settle any more substantive questions about reasons for an attitude. That is to say, it will not distinguish between (what we might call) good and bad reasons—it will not determine which considerations really do bear on a question, or which considerations really do settle a question. The constitutive reasons *bear or are taken to bear* on a question, the answering of which amounts to having the attitude. These are the right kind of reasons for believing. But the fact that a consideration bears or is taken to bear on a question does not yet mean that it is in fact adequate to settle the question. There is room for various sorts of mistake: mistake of fact, mistake about whether the consideration bears on the question, and mistake about whether a consideration that bears on a question settles the question. Further, there is room to distinguish between "reasonable" and "unreasonable" mistakes of each kind. Thus, there remains room to distinguish good constitutive reasons from bad ones, along each of these dimensions.[25]

THE SCOPE OF THE CONSTITUTIVE/EXTRINSIC DISTINCTION

The constitutive/extrinsic distinction can be drawn for any attitude that can be formed or revised simply by settling for oneself a question or set of questions. It seems reasonably clear that belief and intention are such attitudes. Notice that, whenever one has an attitude that can be formed or revised simply by settling for oneself a question or set of questions (regardless of how the attitude was in fact formed), one is committed to an answer to the relevant question(s). One is committed in the sense that, if one has the attitude, one is answerable to certain questions and criticisms—namely, those questions or criticisms that would be answered by the considerations that bear on the relevant question(s). So, for example, if I believe *p*, then I am committed to *p* as true, that is, I am answerable to questions and criticisms that would be answered by the considerations that bear on whether *p*. Likewise, if I intend to ϕ, then I am committed to ϕ-ing, that is, I am answerable to questions and criticisms that would be answered by considerations that bear on whether to ϕ.[26] I call such attitudes "commitment-constituted." The reasons that (are taken to) bear on the relevant question (or set of questions) are the "constitutive" reasons for the attitude, not because the reasons *themselves* constitute the attitude (one may, perhaps, form the attitude for no reason, or may maintain the attitude while having long forgot one's reasons), but because the reasons (are taken to) support the commitment that is constitutive of the attitude.

---

[25] Thus it seems to me likely that investigation into the nature of an attitude, while important for drawing the constitutive/extrinsic distinction, will not settle the more pressing question about the justification of that attitude: Which, among the constitutive reasons, would justify it?

The constitutive/extrinsic distinction can be drawn for any commitment-constituted attitude. For any such attitude there will be constitutive reasons, reasons which bear on the question, the answering of which amounts to forming the attitude. Further, because we are reflective creatures, capable of thinking about our own attitudes, and because which attitudes we have can matter to us in various ways, we could always, in principle, construct a case (perhaps with the use of science fiction) in which certain considerations show the attitude in some way good to have, *without* supporting the commitment constitutive of the attitude. These will be extrinsic reasons. Thus we can, in principle, draw the constitutive/extrinsic distinction for any commitment-constituted attitude.

However, neither the constitutive/extrinsic distinction nor the content-/attitude-related distinction can be drawn for every attitude or every rational activity. Most obviously, these distinctions cannot be drawn for ordinary intentional actions, such as turning left or preparing dinner. Reasons that count in favor of this or that action bear, unambiguously, on the question of whether so to act. One might try to label these "action-related" reasons, but there is no corresponding content-related question. Likewise, one does not perform an ordinary action simply by settling for oneself the question of whether so to act—by settling that question, one therein *intends* to act.[27] There are no constitutive reasons for ordinary action.[28]

---

[26] A point (almost) made by G.E.M. Anscombe, *Intention* (Oxford: Blackwell, 1957).

The distinctions also cannot be drawn for certain propositional attitudes, such as supposing *p* for the sake of argument, imagining there's no heaven, or remembering where you put your keys. There are certainly considerations that bear on whether these are good to do—perhaps it will advance the argument or lead to liberation or allow you to get to the store. One can certainly suppose, imagine, or remember for these reasons. By finding these reasons convincing, one decides or intends or attempts to suppose or imagine or remember. One's attempt can (in principle) fail, as it can for ordinary action. Such propositional attitudes seem to be action-like.[29]

CONSTITUTIVE REASONS AND JUSTIFICATION

I hope to have given an independent way to specify the "right kind" of reasons for commitment-constituted attitudes. We can locate this class of reasons by considering the way in which an attitude can be formed. Commitment-constituted attitudes can be formed simply by answering a question or set of questions. Considerations that (are taken to) bear on the relevant question(s) are the "right kind" of reason for these attitudes.

Still, one might want to ask, why are the right kind of reasons for commitment-constituted attitudes those that (are taken to) bear on a question, the answering of which amounts to forming the attitude? While I do not have a complete answer, we can make a start.

Earlier we saw that the "right kind" of reasons correspond to the "right kind" of justification, and that the "wrong kind" of reasons would provide a corresponding, peculiar sense of justification. We can now see how this is, and why the "right kind" might be right.

---

[27] Preparing dinner is not something one does simply by settling the question of whether to do so. Rather, by settling that question, one therein *intends* to prepare dinner. If all goes well, one will execute that intention in action. But all may not go well, and so the intention and the action remain distinguishable.

[28] An analogous distinction can be drawn in reasons for actions, if the action is picked out according to the reasons for which is performed. Elsewhere I claim that kind, courageous, and spiteful actions, for example, are so discriminated. I explore the analogous distinction, and its importance for moral theory, in "Extrinsic Reasons, Alienation, and Moral Philosophy" (in progress).

[29] Or, they seem to be action-like, in those cases in which they are had for such reasons. In such cases, I am inclined to call them mental acts. I would thus draw the distinction between such "mental acts" and commitment-constituted attitudes by considering whether the attitude can be formed, regardless of its content, simply by settling the question of whether to form it. (This way of drawing the distinction between mental acts and attitudes differs starkly from the common practice of classifying occurrent states as "acts" and standing dispositions as "attitudes.")

Commitment-constituted attitudes, we have said, leave one open to certain questions and criticisms—that is, they leave one open to certain requests for justification and criticisms for lacking the same. These requests and criticisms would be satisfied by adequate constitutive reasons. Thus we have specified, not only a set of reasons, but also a corresponding sense of justification. This is the ordinary sense. Given typical assumptions about agency and justification, it at least seems appropriate that the reasons that justify an activity, in the ordinary sense, should be those that, by finding convincing, one therein partakes in the activity.

Consider a case: whether I am justified in believing the butler did it depends on the adequacy of the reasons that (I take to) bear on whether he did it. By finding such reasons convincing, I will believe. In contrast, if I find convincing the reasons which I take to bear only on whether it is good to believe the butler did it (extrinsic reasons), then I will, therein, form a second-order attitude about the belief that the butler did it. I will believe that it would be good to believe the butler did it, and I may intend to bring it about that I believe the butler did it. Whether I am justified in these second-order attitudes will be determined by the adequacy of the constitutive reasons for *these* attitudes—considerations which (I take to) bear on whether it is good to believe the butler did it, or on whether to bring it about that I believe the butler did it—which are, of course, the extrinsic reasons for the belief that the butler did it. So it seems that the reasons that justify an activity are those that, by finding convincing, one partakes in the activity.

Earlier we saw that, if believing the butler did it would save your life, and if, for that reason, you somehow manage to believe the butler did it, then there is at least a sense in which you are justified and even rational for doing so. Elsewhere I have argued that you cannot form a belief directly for extrinsic reasons. For such reasons you can only act so as to bring it about that you believe.[30] But of course *this* activity—acting so as to bring it about that you believe—might be justified by the reasons that show the belief good to have. Thus, believing that the butler did it can be, in a sense, "justified" by extrinsic reasons, in the same way that sleeping or stumbling or being angry or falling in love might be, in a sense, "justified" by reasons that show it useful to sleep or stumble or be angry or fall in love. In such cases, you have reason to bring it about that you do something which is not itself the kind of thing that can be done by finding those reasons convincing (or perhaps for reasons at all). Believing the butler did it is, in a sense, "justified" by extrinsic reasons only because (or insofar as) *bringing it about that you believe* the butler did it is justified by those reasons in the ordinary way.[31]

---

[30] See "Controlling Attitudes."

THE CURRENT DISCUSSION

Current discussion of the "wrong kind of reasons problem" has focused on reasons for a particular class of attitudes, namely, those "pro-attitudes" such as preferring, desiring, admiring, and even "valuing" which are thought to be appropriate just in case their object is valuable. Current discussion has understood the wrong kind of reasons problem as a threat to "fitting-attitudes" or "neo-sentimentalist" accounts of value.[32]

My ambitions are at once broader and narrower. My ambitions are broader, in that I hope to have solved a broader problem: I hope to have provided a general formula for identifying the "right kind" of reasons for any commitment-constituted attitude (not just for those pro-attitudes involved in valuing). My ambitions are narrower, in that I do not hope to address the metaphysics of value. I will, however, sketch how the account I have provided would apply to the problem under discussion.

The wrong kind of reasons problem is thought to threaten the fitting-attitudes account of value. Such an account attempts to understand value by appeal to whether certain "valuing" attitudes are fitting. (It "passes the buck" from values to attitudes.) Reasons of the wrong kind pose a threat, because they can seem to make an attitude "fitting" without showing its object to be valuable. They thus generate counterexamples.[33] To avoid these counterexamples, we need to identify the reasons that "really" make the attitudes fitting. We need to rule out the wrong kind of reasons.

---

[31] Phillip Pettit points out some subtlety here. In certain cases, one might want to say that the activity of bringing it about that you believe *p* could not be justified at all—because, in the circumstances, that activity is manifestly impossible. Yet, in such cases, certain reasons may well show the belief good to have. Thus, in such cases, one might want to say that the belief that *p* is "justified," in some sense, by extrinsic reasons, while denying that the activity of bringing it about is justified, in the ordinary sense. I imagine this case could be handled in a number of ways. I am inclined to think that the extrinsic reasons, in such a case, "justify" the belief insofar as they are associated with a possible justification of bringing about the belief: you would be justified, if it were possible, in bringing it about that you believe *p*.

[32] Scanlon's "buck-passing" account is characterized in the literature as a "fitting-attitudes" account. I am dubious. I will return to the point in a later footnote.

[33] One such counterexample was noted in an earlier footnote: Rabinowicz and Rønnow-Rasmussen imagine an evil demon who requires you to admire him "on account of his determination to punish [you]" if you do not. The demon seems to have given you a reason to admire him, a reason that makes admiration, in some

I hope to have provided a way to identify the right kind of reasons for attitudes. If I am right, then, in order to determine the right kind of reasons for the vast and interesting range of attitudes with which the current discussion is concerned (attitudes such as admiring, preferring, fearing, envying, valuing, and desiring), one will have to determine the commitment(s) implicit in each. That is, one will have to identify the question or set of questions, the settling of which amounts to having the attitude.[34] The right kind of reasons for these attitudes will be those considerations that (are taken to) bear on those questions.

One might, then, be doubtful of my proposal or of its applicability to these attitudes, because one might doubt that these attitudes are commitment-constituted—one might doubt that they can be formed by settling for oneself a question or set of questions. But I suspect they can, on the basis of the following reflection:

Certain attitudes are subject to direct rational criticism: one can be asked to give one's reasons for them. One can be asked why one believes, intends, resents, supposes, or imagines, where this question does not look for an account of *how it came about that* one believes, and so on, or even why *one brought it about that* one admires, and so on, but rather for the reasons for which one might have formed the attitude. Such requests for justification can, in principle, be satisfied, and it seems that one can satisfy such requests in one of two ways: by providing reasons that show the attitude in some way good to have or by providing reasons that bear on some other question(s), the settling of which amounts to having the attitude. If one can satisfy the request for justification simply by showing the attitude good to have, then it seems the attitude is action-like (like remembering where you left your keys). If you can satisfy the request for justification by

---

sense, fitting, and yet it is a reason of the wrong sort to show him admirable. See "Strike of the Demon," p. 414.

[34] The claim that settling a set of questions "amounts to" having the attitude may seem a stronger claim than it is. Settling the question of whether p amounts to believing p in that, if one settles the question, one is rightly understood as believing, and, if one believes, one is rightly understood to have settled the question—as shown by the fact that one believes just in case one is vulnerable to questions and criticisms which would be satisfied by considerations that bear positively on the question. Likewise, if settling some complex set of questions "amounts to" resenting, say, then, if one settles the set, one is rightly understood as resenting, and if one resents, one is rightly understood to have settled that set of questions—as shown by the fact that one believes just in case one is vulnerable to questions and criticisms that would be satisfied by considerations that bear positively on the set of questions. I consider these connections briefly in "The Will as Reason" (in progress). I am grateful to Mark Johnston, Barry Lam, Victoria McGeer, and Michael Smith for pressing on this claim, though I am confident I have not fully addressed their concerns.

providing reasons that bear on a question, the settling of which amounts to having the attitude, then the attitude is commitment-constituted. Thus, it seems that, insofar as an attitude is itself capable of such direct rational criticism, either it will be action-like or it will be commitment-constituted. The pro-attitudes involved in valuing seem capable of direct rational criticism (I can be asked why I admire or prefer, not merely how it came about or why I have brought it about that I admire or prefer). So it seems that these attitudes are either action-like or commitment-constituted. But if an attitude is action-like, there will be no distinction between constitutive and extrinsic or content- and attitude-related reasons for it. The wrong-kind-of-reason problem, itself, suggests that there is some such distinction. Thus, it seems likely that, insofar as the problem can arise for these attitudes, they are, in fact, commitment-constituted.[35] And thus I am hopeful that, by doing at bit of philosophical excavation, we can come to understand the commitments implicit in them. We would thus isolate the right kind of reasons for them—they are the reasons that support those commitments.[36] (Of course, identifying the commitments implicit in such attitudes will be a difficult and delicate task, one we will likely have to approach piecemeal.[37])

Thus I am hopeful that we can eventually distinguish the "right" from the "wrong" kind of reasons for those attitudes involved in valuing. However, despite my optimism on this score, I doubt that identifying the right kind of reasons for these attitudes will be of much help in providing an account of value—because, as we have seen, the constitutive/extrinsic distinction does not itself isolate "good" constitutive reasons for an attitude from "bad" ones. (Even if we could identify the questions the settling of which amounts to admiring or cherishing something, we would not thereby know whether those questions ought to be answered positively in a given case. Merely identifying the right kind of reason for valuing attitudes will not identify the objects of value.) So, if one hopes to give a general account of value by appeal to the appropriate reasons for an independently identifiable class of attitudes, problems will remain. Thus, even with an answer to the wrong kind of reasons problem, we may not be able to pass the buck from value to attitudes.[38]

---

[35] Here is another, perhaps stronger, reason to think that the attitudes involved in valuing could not be action-like: anything action-like is voluntary—that is, it is something one can do simply by posing and answering the question of whether to do it. An attitude that is voluntary in this way could not express any conviction that the object is of value.

[36] D'Arms and Jacobson talk about the "norms of fittingness" for different attitudes. See "The Moralistic Fallacy: On the 'Appropriateness' of Emotion," *Philosophy and Phenomenological Research*, LXI (2000): 65–90, and "Sentiment and Value." They express their doubts about what they call "quasi-judgmentalist" accounts of emotion in Anthony Hatzimoysis, ed., "The Significance of Recalcitrant Emotion (or, Anti-Quasijudgmentalism)," in *Philosophy and the Emotions* (New York: Cambridge, 2003), pp. 127–45.

[37] Elsewhere I have considered whether resentment might be understood as the settling of a complex set of questions—about the genuineness of the wrong, the status of the wrongdoer, one's own status, and the ongoing significance of the wrong done. Settling these questions would be constitutive of resentment, and reasons that bear on them would be constitutive reasons for resenting. Other reasons, which show resentment useful or costly or inconvenient, would be extrinsic reasons for or against resenting. Forgiveness and excuse, I have suggested, should concern the constitutive reasons for resentment. Excusing involves changing one's answer to one of the first three questions. In so doing, one undermines resentment. Forgiveness could be understood as overcoming resentment by revising one's answer to the last of these questions. In contrast, the fact that resentment is a costly and destructive emotion counts in favor of overcoming or eliminating it, but is an extrinsic reason for its elimination; one could not excuse or forgive for such a reason. Rather, one might undertake some sort of drug treatment aimed at eliminating the emotion. See "Articu-

CONCLUSION

I have offered a suggestion about what a reason is. A reason, I have suggested, is a consideration that bears on a question. The basis for this suggestion is the elementary reflection that a reason is an item in a piece of reasoning. (A good reason is an item in a sound piece of reasoning.) In its support, I have shown one advantage over the thought that a reason is a consideration that counts in favor of an action or attitude:[39] this formulation generates a thoroughgoing ambiguity in reasons for certain attitudes—we cannot distinguish precisely between "content-related" and "attitude-related" reasons. If we instead take a reason to be a consideration that bears on a question, we can distinguish between questions, thus distinguishing these classes. The attitude-related reasons count in favor of the attitude by bearing on

---

lating an Uncompromising Forgiveness," *Philosophy and Phenomenological Research*, LXII, 3 (2001): 529–55, and "The Force and Fairness of Blame," *Philosophical Perspectives*, XVIII, 1 (2004): 115–48.

[38] It is well worth noting that this is not Scanlon's proposal. Rather, Scanlon proposes "passing the buck" from value to *reasons*. That is, he maintains that *reason* is a more fundamental normative notion than *value*, in that claims about value could, in principle, always be more illuminatingly put as claims about reasons for certain activities—reasons provided by features of the valuable object or state of affairs other than its "value" or "goodness." It seems to me that maintaining *this* buck-passing thesis does not require that one be able to specify, in any general way, exactly which attitudes are characteristic of valuing, or what the right kind of reasons for them are. The ability to specify which attitudes, in general, are involved in valuing and what the right kind of reasons for them are would be required only if one were hoping to reduce claims about value to claims about certain attitudes—if one were hoping to pass the buck from value to psychology. But that is not Scanlon's suggestion. See *What We Owe to Each Other*, pp. 95–100.

[39] In "Reasons for Action," I consider another advantage.

whether the attitude is in some way good to have; the content-related reasons count in favor of the attitude by bearing on some other question. I hope to have shown, however, that this distinction will not isolate the "right kind" of reasons from the wrong kind, because this distinction is not exclusive. The proposed formulation helps to distinguish the right from the wrong kind of reasons by requiring us to consider the relation between the question on which a reason bears and the attitude of which it counts in favor. The constitutive reasons are those considerations that (are taken to) bear on a question or set of questions, the settling of which amounts to having the attitude. The remaining reasons are extrinsic. Drawing this distinction solves "the wrong kind of reasons" problem; it allows us to distinguish the "right kind" of reasons for certain attitudes—the constitutive reasons—from the merely extrinsic ones.

PAMELA HIERONYMI

University of California, Los Angeles

DECEPTION UNRAVELED*

When deceiving a person is wrong, what makes it so? In the middle of *Truth and Truthfulness,* Bernard Williams answers this question twice.[1] He argues that the wrong in deception should be understood in terms of a deceiver breaching a trust with the deceived. And he argues that it should be understood in terms of the deceiver manipulating the deceived. Williams offers these accounts contiguously within the space of a long book, but he does not explain their relationship: whether there is any tension between them, or whether instead the two accounts are consistent, complementary, perhaps even stylistic variants on a single idea. This relationship seems puzzling in a way worth study. I will suggest that the breach of trust and manipulation accounts of deception are related directly: breach of trust is a kind of manipulation. I will also maintain that not all manipulation in deception involves a breach of trust, and that deception that involves a breach of trust may involve a wrong that is distinguishable from that which occurs in other deception. Deceiving is often best understood as a defensive act governed by the norms of self-defense. Deceiving that involves breach of trust and deceiving that involves other forms of manipulation are acceptable under very different circumstances, I will conclude.

Because Williams does little to distinguish between the ideas of breach of trust and manipulation, it will not suit my purposes to rely on him for an account of the ideas. This presents a difficulty: there is no standard or uncontroversial account of manipulation or trust, and no adequate account of the ideas can be completed in this brief paper. I will therefore develop accounts of manipulation and breach of trust to the extent necessary to make my argument about deception. I will first present cases that I hope elicit tensions in judgment about the acceptability of deception, and then explore how ideas of breach of trust and manipulation more generally may help explain these tensions.

Many of the cases that I discuss will be taken from commercial life. There are two related reasons for this. First, what makes deception

---

*For helpful discussion, I am especially indebted to David Wasserman. I also thank Thomas Dunfee, Nien-hê Hsieh, Howard Lesnick, Mark Liberman, Maurice Schweitzer, David Silver, William Tyson, Scott Weinstein, and audiences at the University of Maryland CP4 Workshop, and the Wharton Legal Studies Department.

[1] (Princeton: University Press, 2002), chapter 5.

0022-362X/05/0209/458–73

often seem acceptable are strategic factors, for example, the need to protect private or proprietary information; these factors are comparatively accessible in commercial contexts. Second, the leading philosophical discussions often concern deception about bewildering problems of life and death, love and betrayal. Yet morality matters not only in such matters of high stakes but also in the mundane choices that we face every day. Commercial life promises cases of deception that are emotionally cool enough to limit the risk of confounding but rich enough to bring out the structure of deception, I suggest.

MECHANISMS FOR SECURING BELIEF

You are thinking about buying a used car from me. You live in the desert, where you take long drives on lonely roads. So you want a car with a cooling system that works well. You ask me about the condition of my car's radiator. I tell you that a mechanic inspected the car this morning, and that he said that the radiator is fine. I know more, but do not say: the radiator would leak if I had not poured Gunk in it to temporarily plug a hole. I brought the car to this mechanic because he was incompetent, would not notice the Gunk, and hence would provide good cover for me. When I told you that the mechanic found no problem with the radiator, did I wrong you? I did not lie. By repeating, without comment, the words of the mechanic, I created an impression that there was no reason to doubt what he said. I wrongly deceived you. I told a truth, but misled you. What more can be said about the wrong that I committed? By deceiving you, did I, as Williams might suggest, wrongly manipulate you? Or breach your trust?

On the surface, the charge of wrongful manipulation seems plausible. I take it as a sufficient even though not exhaustive characterization of manipulation that one person manipulates another when he intentionally causes that person to behave as he wishes through a chain of events that has the desired effect only because the manipulated person is unaware of that chain.[2] "Car Sale" seems manipulative according to this characterization, because in it I cause you to have a false belief in order to get you to behave as I wish, and in doing so, I willfully hide from you facts about the causal process, including the fact that the mechanic was tricked. Also, I seem to breach a trust: by my behavior, I seem to vouch for the mechanic's assessment, inviting your trust and then breaching it. There is perhaps a strong tie between breach of trust and manipulation in "Car Sale." After all, my success in manipulating you seems deeply bound with my success in getting you to trust what I convey about my experience with my mechanic.

---

[2] I acknowledge that other kinds of manipulation, for example, that which occurs by preying on a person's emotions, may be transparent to their victims. But these are not my concern here.

Soon I will argue that in assessing the wrongness of a particular deceptive act, it is important to distinguish between breach of trust and other manipulation elements. In cases such as "Car Sale," the distinction may appear less than clear. But it does not take much changing of the case to sharpen this appearance. One might do so, for example, by supposing that I cultivate trust by solemnly explaining that for the last ten generations, my family's deepest commitment has been to Quaker values according to which it is more important to be honest and forthcoming, and to act for the sake of others, than to make money. My pronouncements—true but misleading, because they represent my family's past but not my present intentions—are a solicitation of trust, and may play a distinctive causal role in eliciting your confidence in what I say or suggest. By making these pronouncements, I convey that you can rely on my goodwill, that is, my intention to act for your sake and not simply for my advantage.[3] I manipulate you, because in ways that I willfully hide from you, I cause you to have a false belief about whether I have goodwill toward you, and then exploit that belief in order to get you to behave as I wish. But because I emphasize the goodwill inherent in my Quaker background, trust plays a distinctive role in this manipulation.

Anyone who has glimpsed the recent large literature on trust will notice that not all would accept my characterization of trust in terms of goodwill. Sometimes, one might object, you trust another person to be truthful not because of his goodwill, but because you know that it is in his interest to advance your interest, and he has reason to think that his being truthful will advance your interest.[4] For example, a repeat buyer of fish may believe the fishmonger's report of freshness because he knows that it is in the fishmonger's interest to retain him as a future customer, and not because the fishmonger has goodwill in a stronger sense. The fishmonger, as I imagine him, exhibits no goodwill, because he does not act for the sake of his customer, but instead acts in order to get an advantage for himself. No doubt the sort of reliability that the fishmonger demonstrates is socially important. At least for purposes of this paper, his reliability is best characterized without using the idea of trust. I propose that one may regard someone as reliably truthful because of his goodwill, and one may be regard him as reliably truthful because of personal advantage he gains from truthfulness; there is reason, I will suggest, for reserving the term 'trust' for the first kind of regard, but by doing so not denying the importance of the second.

---

[3] The idea that trusting a person requires viewing that person as possessing relevant goodwill is ably advanced in Annette Baier, *Moral Prejudices* (Cambridge: Harvard, 1995), chapter 6.

[4] For representative examples of this approach, see Elinor Ostrom and James Walker, eds., *Trust and Reciprocity* (New York: Sage, 2003).

One need not demonstrate trustworthiness to secure credibility. One can, for example, secure it externally, by providing evidence that one has reason for being truthful even though the reason functions independently of one's goodwill. One can intelligibly say to another: "I know that you don't trust me, but you nonetheless have reason to believe what I say; if what I say is false, I will suffer the consequences." If a mafia don tells me that I must be truthful to you and publicly threatens me if I am not, there is reason to think that I will be truthful, and it has nothing to do with my trustworthiness. Indeed, if it is known that were it not for the mafia don, I would lie through my teeth, that need not undermine my reliability as a truth teller, so long as the mafia don is known to remain consistent with his threats. Thus there is good reason to distinguish between a trustworthy truth teller and a reliable truth teller.

Despite the artificiality of "Mafia Don," there are similarities between it and common cases in which one secures credibility without relying on trust. In "Car Sale," for example, one bolsters one's credibility by finding reasons to believe that what one suggests is corroborated by an independent, external source—a mechanic. Here, it seems that you believe what I suggest (that there is no problem with the radiator) not because of trust, but because you learn about a report by an external source. Perhaps you trusted the mechanic, but he did not breach a trust with you. And you did not trust me; you would never have believed me without the mechanic's report. So I manipulated you, and deceived you, but none of this involved my breach of trust.

Both the mafia don and the auto mechanic, in different ways, enhance one's credibility without appeal to trust. One may find more mundane ways in negotiation to establish credibility in the absence of trustworthiness. For example, one might reveal evidence about the meagerness of one's assets or about the attractiveness of rival offers, hoping to show that one has a good reason for having a particular settlement preference. The existence of this evidence would not show that one is trustworthy, but it might show that there is good reason to think that one is reliably truthful about the stinginess of one's settlement preference.

Just as one need not rely on the appearance of trustworthiness to attain credibility for what one says or suggests, one need not rely on this appearance in order to deceive and manipulate one's audience. Suppose that I create the illusion of a mafia don in order to bolster my credibility. Beacuse the don is my creation, I manipulate my audience in using his story, just as I would if I had used another form of manipulation and solicited the audience's trust. In each case, I cause my audience to have false belief through a process that succeeds only because I make it opaque to that audience. In one case, the manipulative process is in gaining trust about my truthfulness, and then breaching it; in the other, it is in causing you to believe that there is an external basis for my truthfulness, and then using that credibility to cause you to have a false belief. There is, thus, substantial variation in manipulation. One can deceive a person in a way that manipulates him but does not breach his trust.

So far, I have tried to sketch two distinguishable processes through which deception occurs: manipulation that involves a deceiver breaching trust and simple manipulation, which does not rely on the idea of trust in this way. In the next section I will argue for the moral acceptability of certain deception that does not rely on trust. I aim then to show how adding an element of trust makes a difference for the moral assessment of deceptive acts.

UNAVOIDABLE DECEPTION

However wrong my misleading statement seem in "Car Sale," there are other cases in which misleading statements seem less troubling, despite the fact that they involve intentionally causing a false belief in the mind of an innocent person. In this section I will consider another example from negotiation, arguing that certain deceptive acts that occur in negotiation are morally acceptable because of their defensive function. To do this, I will develop an analogy between, on the one hand, the use of force to defend oneself against an act of physical harm and, on the other hand, the use of deception to defend oneself against a query that risks economic harm. I will maintain that a straightforward application of traditional principles of self-defense justify the use of deception in common negotiation.

Consider "House Sale." You ask me whether I would pay as much as $300,000 for your house. My voice dripping with doubt, I say, "I might go that high; but it is an awful lot of money for me." What I say is literally true, even though I thought that I would have to pay much more than $300,000, and would in fact have been happy to pay $400,000. My statement that I might go that high is misleading and deceptive, because it falsely suggests something about my settlement preferences—that I might not be happy, or even willing, to pay $300,000. But notice that by misleading you in this way, I need not breach your trust. You regard me as an adversary, someone trying to get the best of you. Hence you expect me to give you little more than the literal truth in our discussion. In this respect, you can expect no more from me than our legal system expects from a hostile witness being cross-examined by an attorney in a trial: the witness may not lie, without committing perjury; but as long as what the witness says is literally true, he is not responsible for false suggestions arising from his statements. Similarly, I will maintain, in a competitive negotiation, as long as one remains in the realm of settlement preferences, one is not responsible for more than the literal truth of what one says; one is permitted to defend oneself by making false suggestions.

"House Sale" and "Car Sale" are not morally on par, but they involve important similarities. In both cases, a party deceives in order to hide information that his bargaining foe would find useful. Both cases therefore involve protecting oneself against possibly damaging invasive queries. Even so, the two cases differ in a way that renders salient some limits on the right to use self-defense. Hence I will argue for the moral acceptability of deception about settlement preferences in "House Sale," and then distinguish it from the deception in "Car Sale." My argument will draw on traditional ideas about the norms of self-defense.

Consider a case of self-defense from outside the realm of deception. I credibly threaten to slug you. May you use force to prevent my imminent attack? It depends. One may use violence defensively only if it is unavoidable, that is, only if one cannot otherwise, through less costly means, prevent an attack. Call this the *Avoidability Principle.* If there is a policeman nearby whom I do not see but you know could, through nonviolent intimidation, secure the peace, then the Avoidability Principle requires that you ask for his help rather than use force. The Avoidability Principle, thus described, applies to the use of violence in self-defense. We can easily extend it so that it applies to the use of deception in self-defense, saying simply: "One may use violence or deception defensively only if it is unavoidable, that is, only if one cannot otherwise, through less costly means, prevent an attack." In its defensive use, deception seems often to satisfy this principle. Indeed, the idea of unavoidability is part of the most discussed example—at least among philosophers—of acceptable deception: a murderer asks you for the location of his intended victim, who happens to be hiding in your cellar, and you contemplate deceiving him. Few doubt that it is morally acceptable to deceive in "Murderer." By deceiving in this case, one defends an innocent person against a wrongful threat, and uses the least objectionable means possible to do so.

"House Sale" is not "Murderer." Nothing profound is at stake. Life goes on if one does not deceive. Indeed, there are several ways to avoid deceiving: one might disclose one's pricing preferences, or refuse to answer queries aimed at eliciting those preferences, or simply walk away from the deal. Because the deception in "House Sale" seems easy to avoid and hence in some sense avoidable, the deception may appear not to satisfy the conditions for legitimate self-defense, and thus appear morally dubious. These appearances, I will argue, are illusory: available paths of avoidance have costs that one should not be expected to bear.

There are limits on what one must do to avoid defending oneself, even outside the realm of deceptive defense. In the traditional jurisprudence of self-defense these limits may be understood in terms of the duty to retreat. Thus if someone threatens to attack a person on that person's property, and the prospective victim can safely avoid the attack by retreating into his home, then he has no duty to retreat rather than to fend off the attack using reasonable force; when he stands on the right to not retreat, he may use reasonable force to protect himself. In this case, the use of force remains unavoidable, because one has available no reasonable means to avoid it while standing on one's right. Deception in negotiation sometimes seems to exhibit a similar character. Even though one might be able to avoid deception by walking away from a deal, or taking the loss that comes from exposing one's reservation price, one has no duty to absorb these costs, and so one may stand his ground by deceiving the other party about, for example, one's settlement preferences, I will maintain.

My argument about the unavoidability of defensive deception in negotiation becomes easier if I make what I take to be a fairly uncontroversial assumption: any party in a competitive negotiation has reason to learn what his bargaining foe would regard as his worst acceptable price, that is, his reservation price. This assumption is plausible because information about your opponent's reservation price is normally financially advantageous in negotiation; it gives you a basis for knowing how hard to press for concessions, and for knowing when pressing for concessions makes no sense. Information about reservation price is particularly useful in a case like "House Sale," in which there is nothing but price about which to bicker. In a competitive negotiation, every gain in price that your foe gets is a loss for you. Thus because your foe gains the prospect of a financial advantage by learning your reservation price, you have reason to protect the privacy of information about that price. Now suppose that your bargaining foe directly asks about your reservation price. It is against your interest to reveal the truth, and you have no obligation to incur the cost that would come from doing so. What are your options? You might refuse to answer. But then, your bargaining foe, seeking the only information that might help him get a good price, has reason to not rest satisfied with your refusal. He might thus prod for your price by sending trial offers and asking you for explanations of your response to these offers, in the hope of gaining the information to make an intelligent guess about your position. If you try to stay clean of this process by maintaining silence in the face of these forays, this may send its own message, for example, that the truth is too embarrassing to speak; silence may seem like a strategic ploy. No matter what you say or do, your opponent may use it as a basis for making an inference about your reservation price. You have no reason to acquiesce in his reading your behavior (and thus using your behavior) in ways that may be unfavorable to your interests. It follows that the prudent course in negotiation is to defend yourself against this prospect of harmful curiosity by transmitting misleading messages about your reservation price. Because of the normal structure of competitive negotiations like "House Sale," deception about reservation prices may thus be generally unavoidable, if you want to stand your ground and protect your bargaining position.

I have now tried to make plain how the norms of self-defense may be used to vindicate deception that aims at fending off the prospect of economic harm. My focus has been on answering skepticism that such deception can satisfy a traditional requirement for the use of self-defense, the requirement of unavoidability. If my argument thus far is correct, it leaves open the question how much deception it licenses. This is a question I take up in the next two sections.

LIMITS OF UNAVOIDABILITY

It may seem unsettling that, as I conclude in the previous section, one may sometimes deceive to protect against mere economic loss. One might worry that the standard assumed seems promiscuous. It might seem to sanction, for example, the deception about my radiator in "Car Sale," because that deception may seem unavoidable if I am to protect my price. Indeed, the standard might seem to sanction much of what the law deems fraud. The worry is misplaced, however, because I never suggested that deception is relevantly unavoidable whenever it is the most effective means for protecting economic advantage. In fact, in the cases of the sort that we have so far considered, you may deceive only to protect proprietary information, information about which your audience has no right to know. In general, you cannot plausibly claim unavoidability for defensive action that serves the purpose of wronging a person, whether the defensive action takes the form of deception or force. If I want to invade your house to liberate a kidnap victim, and the only way you can fend off my attack is by using force against me, then your use of force is not unavoidable. You have no right to deceive in order to do wrong. Undermining a person's choice about the fundamental nature of what he buys, merely to gain an economic advantage over him, is wrong. So the deception in "Car Sale" is wrong. But this may raise a new worry. If it matters, in determining unavoidability and thus in vindicating deception, whether your audience has a right to know, then perhaps it will suffice for all casuistic purposes to focus simply on whether the audience has that right. What is the relevance of self-defense?

I wish to show that self-defense has ineliminable relevance in assessing deception. I will argue that the casuistry of self-defense constrains how and when one may deceive in ways not reducible to considerations about the audience's right to know. To do this, we must find a case in which we hold constant the moral situation of the deceived, but vary the situation of the deceiver only in ways that directly alter the acceptability of deception. Here, again, it will be useful to draw on traditional norms of self-defense, in particular, the principle that if you are responsible for creating your own need to use self-defense, then that limits or eliminates your right to use self-defense. Suppose, for example, that one evening, I enter a dangerous neighborhood, because I am lost. Hoodlums threaten me, and I run. My pride wounded, I return the next night, carrying a gun, hoping that the hoodlums again threaten me, so that I might shoot them in self-defense. In this case, because I gratuitously create the need to use self-defense, I undercut my claim to the right. I have a right to self-defense, but not a right to use force to protect against trouble that I have sought out. More generally, a person has no right to claim that self-defense is unavoidable when he bears some responsibility for the attack he faces.

The relevance of this limit on the right to self-defense can be seen in a recurring problem of deception in securities law. The problem involves cases in which corporate officers are asked about arguably proprietary information that would, if disclosed, adversely affect the value of their firm's stock. Suppose that a firm's viability is staked to the successful development of a particular product. And suppose that there has been a setback in its development. You, a corporate officer, are optimistic that the product will eventually be a success, but reasonable people might disagree. If word gets out about the difficulties, the value of the firm's stock might be so adversely affected that it would be unable to develop the product. What should you do when journalists and your shareholders ask about progress on the product? If you disclose, the harm is clear. If you say no comment, you encourage adverse inferences that will likely have the same harmful effect because people will wonder what you are hiding. When facing these problems, many firms find ways to say things that are literally true but misleading; they deceive in order to protect what they regard as proprietary information. Is this deception acceptable?

Courts disagree about how to handle this problem. Some tolerate misleading statements; others do not. Judge Richard Posner takes the tolerant approach in *Eisenstadt v. Centel,*[5] an influential case in which the proprietary information was not about product development but about an unsuccessful auction. If word got out about the auction, it would depress stock prices. So corporate officers faced a sensitive problem when journalists and members of the public asked about the success of the auction. Centel's most troubling comment, in response to these queries, was to say that the bidding process went well.[6] In fact, only a handful of parties bid, and none came close to an acceptable price. The auction was a bust.

I find cogent the case that Centel wrongly deceived its stockholders. Posner does not. He maintains that Centel's comment that the bidding process went well was true, because "the auction process went as smoothly as could be desired."[7] By that, he means merely that the auction opened in an orderly way, bids were taken, assessed, and so forth. Substantively, Posner acknowledges, the auction "failed," because it produced very few bids and no remotely acceptable bid. Since the auction failed, was Centel's statement that the process went well misleading, if not false? Posner thinks not, because he believes that it was plain that the intended audience would not form a false belief on the basis of Centel's statement about the auction process. Posner's assertion about audience response to Centel's statement is, of course, pure conjecture; he gives no evidence for it. Worse, it relies on mystery. If it is as obvious as Posner maintains that nobody would believe Centel's statement and suggestions, then why should corporate officers bother to make them? Why say or suggest something that you know your audience will not believe? Posner's answer to this question invokes an analogy with writing letters of recommendation in an environment in which inflation, or exaggeration of praise, is the norm. If it is the community norm to say that merely good candidates are wonderful candidates, then if one says that one's own merely good candidate is (merely) good, one's words, or what one suggests by uttering those words, will be understood according to the community norm, and one will send a false message about the candidate, Posner thinks.

---

[5] 113 F. 3d 738 (1997).

[6] At 738. In the opinion Posner quotes Centel officers as saying that the "bidding process" went well. When Posner discusses what officers said, he instead speaks about the "auction process" going well. I will follow Posner in ignoring any difference in the meaning of these phrases.

[7] At 746.

By invoking the letter-writing analogy, Posner suggests that saying that the auction process went well is merely a complex way of conveying some minimal truth about the auction, perhaps that the process started and ended smoothly. But the letter-writing analogy is inapt. In its statements about bidding, Centel did more than exaggerate an evaluation. It encouraged false belief about the occurrence of specific events. Contrary to Posner's benign view, there are plausible scenarios in which even a reasonable person would form false beliefs on the basis of Centel's statement that the bidding *process* went well, and conclude that acceptable bids were made. A reasonable person might, for example, not notice that Centel praised the bidding process, rather than the bids themselves. More important, even if a reasonable person did notice the word "process," he might still quite plausibly believe that some acceptable bids were made, because he might reasonably think that part of a successful process is that acceptable bids are made. It is easy to imagine other statements about process that, like Centel's, are misleading. Suppose that a scientist says that the experimental process goes well; his justification is that he successfully collected and tested all needed data; but he does not mention the data disconfirm his hypothesis. Or suppose that a lawyer claims that the trial process goes well; his justification is that all witnesses show up, the proceeding begins and ends in a timely way, and procedural formalities are respected; but he does not mention that he is decisively losing. The scientist's and the lawyer's attempt to hide behind a technical idea of pure process are ruses aimed at hiding failure. So is Centel's.

The best argument for Centel cannot rely on Posner's idea that its officers said nothing false or misleading. Instead that argument must establish something that a judge might understandably be reluctant to say: that it was morally acceptable for Centel to make a false or misleading statement. In trying to make that argument, the similarity between "Centel" and "House Sale" is crucial. In "House Sale," it was morally acceptable to make misleading statements for defensive reasons: in order to fend off queries about matters that, if disclosed, might prove harmful, and about which one's audience had no right to know. Centel's best argument must make the case that it was similarly acceptable for Centel to make misleading statements that would fend off queries about matters that, if disclosed, might prove harmful, and about which the audience had no right to know. There is a problem for such an argument. "Centel" and "House Sale" differ in a crucial respect that bears on the issue of the avoidability of deception. Centel officers, unlike the house seller, could have ex ante adopted a policy that would allow them to avoid the need to deceive. Suppose that Centel announced a policy before the auction: we will hold an auction, but we will make no comment about it until one month after the auction ends. Then there would have been no need to make a misleading statement about the auction process because Centel's silence would not be subject to the adverse interpretation that the truth was too ugly to speak. This is an option more readily available to Centel than to the house seller because unlike Centel, the house seller is not an institution with the capacity to make public commitments in advance of a transaction. But Centel chose not to make that public commitment, knowing that the intrigue of an auction might strengthen stock prices. By choosing to stay silent, Centel created the conditions that made it necessary for it to defend itself from queries about the auction. Because it was responsible for its need to deceive, it forfeited the right to use deception defensively, and wrongly deceived people about its auction.

In this section, I have used the example of Centel to argue that norms of self-defense are important in assessing deception, and not reducible to considerations such as an audience's right to know. Some deception is morally acceptable, I have maintained, because it provides an unavoidable defense against invasive queries that threaten harm. I have also argued that Centel's deceptive statements do not satisfy the unavoidability condition, and were therefore wrong. Centel's predicament was not unique. As I suggested, it is shared by many firms that face the prospect of disclosing embarrassing news. One may, of course, disagree with me about how a firm should respond to this predicament. If my argument so far is correct, then the best way of doing so acknowledges the relevance of norms of self-defense, arguing that an analogy with "House Sale" is strong: that because that there is no feasible strategy of pre-commitment to silence available, the right of self-defense warrants making misleading statements.

DECEPTIVE WRONGS

No doubt some, taking a position that I will call *Moral Purism,* would react differently to the prospect of intrusive queries in competitive contexts like "House Sale" and "Centel," saying that because deception is inherently wrong, no matter how much it might cost to forswear it, one should not deceive. In important respects, Moral Purism resembles *Pacifism,* which finds wrong in all intentional violence. Something is true in Pacifism. Intentionally harming an unwilling human being, even in pursuit of the social good, seems abhorrent. Despite this, Pacifism seems an unreasonably rigid view that ultimately tolerates too much violence. Similarly, there may be truth in Purism, even though its scope is limited.

Consider another version of "House Sale," "Quaker House Sale." This time, I, the buyer, employ the Quaker Gambit. I invite you to trust me, suggesting, but not stating, that because of my distinctive faith, I would not deceive you about anything, including my settlement preferences. Then, by uttering an effective half-truth, I deceive you about my settlement preference. I cultivate your trust and breach it, just so I might gain an economic advantage. Here my behavior, because it breaches trust, seems deeply and unequivocally wrong. That is the truth Purism sees: deception that breaches a trust is typically a clear wrong. The problem in Purism is that it overgeneralizes this truth, assuming that the wrong that occurs in deception by breach of trust is matched by the magnitude and kind of wrong that occurs when one deceives through simple manipulation. I will maintain instead that deception that relies on breach of trust is comparatively problematic deception. More specifically, I will maintain that deception that relies on breach of trust is always wrong when used in pursuit of economic gain, even though it may be acceptable when used against a substantial wrong.

As we have been considering cases in which deception might be rendered acceptable by virtue of its role as a defensive device, it will be useful to consider again more general ideas in the traditional ethics of defense. One element provides that any defensive measure employed must be proportional to the threat that one faces. Hence it is wrong to defend oneself against the threat that someone will steal one's coffee by blowing off his head with a bazooka, even if that is the only effective protective measure one has available. In commercial interactions of the sort that we have been considering, I suggest, a breach of trust is also disproportionate to the threat it would aim to thwart. A breach of trust is not an attack of the same magnitude as a bazooka attack, but it is a more substantial invasion than that which occurs when one deceives without breaching a trust.

How does deception by breach of trust compromise autonomy? "Quaker House Sale" provides an example. Suppose that you are seduced by the Quaker Gambit, and so trust me on the matter of my settlement preferences. You cannot entertain skepticism about these preferences. Indeed, it makes no sense to say that you trust what I say about my settlement preferences, but that you at the same time entertain skepticism about what I say. Trusting me about what I say regarding my settlement preferences means nothing if not that you accept what I say as true, without skepticism. Because you accept what

I say as true, you have transferred the effective locus of your decision-making on your beliefs about my settlement preferences to me. Your autonomy with respect to this matter is in my hands. Thus if I deceive you and betray your trust, I compromise your autonomy.

There is a natural objection to the above line of argument. One may complain that autonomy cannot be compromised so easily. If you choose to accept what I say about my settlement preferences in "Quaker House Sale," and what I say turns out false, that has no bearing on the autonomy of your choice. Instead, by believing me, you autonomously took a chance, and lost your gamble, just as you might in coming to believe any other uncertain proposition, one might argue. When you believe me, you do not get the truth, and hence do not get what you want. But plainly, simply failing to give a person what he wants is not thwarting his autonomy. Deceiving a person who trusts you about your settlement preferences, one might thus think, may cause him to make a mistake, but does not undermine his autonomy.

This objection misunderstands the constraints imposed by respect for autonomy. It supposes that if you autonomously choose to hand over decision-making authority to a person, as you do when you genuinely trust him, and that person then exercises the authority, your autonomy is thereby not abused or otherwise compromised. But the mere fact that you autonomously choose to hand over your decision-making authority does not show that your autonomy is respected when that authority is exercised. Showing respect would require that the authority be exercised in a way that is consistent with your reasons for handing it over. Still, the objection may seem hard to assess, because it is hard to say in general terms what constitutes compromising autonomy and what constitutes respecting it. Perhaps it will be easier to start from the particular, a clear case of undermining autonomy after authority has been handed over.

Consider this case. Hannah expects to be unconscious when she soon undergoes surgery. It is possible that difficult life and death decisions will have to be made at that time, but the options cannot be enumerated in advance. She has discussed her values about such matters with Anna, and so she writes a note designating Anna as a surrogate decision-maker. The surgery does not go well, Anna is consulted, and knowingly chooses against Hanna's values. It seems clear that Anna compromises Hanna's autonomy. One might argue: because Hanna was able to exercise her choice that Anna acts as her surrogate, her autonomy was respected. But that seems implausible because Anna was entrusted to act consistently with Hanna's values and she did not. Anna breached a trust, acting in a way that is presumptively wrong, and that might be excusable only in the most extraordinary circumstances. For morally similar reasons, I suggest, deception through breach of trust seems presumptively wrong, nothing that can be justified or excused by the mere prospect of the pecuniary gain that drives typical negotiation. Breaching trust is a grave act, an assault on autonomy, something that we might entertain in response to a grave wrong, such as the threat of murder, not something that we should entertain merely as a way to gain advantage in a competitive negotiation.

If I am correct about the wrongness of deception based on a breach of trust, it does not follow that all deception is wrong. Not all deception involves a breach of trust. Recall "House Sale."

When I complain that $300,000 would be a burden for me to pay, you remain skeptical. You know that I am trying to make plausible that my reservation price is low. So you take my comment as one piece of evidence, to be weighed with other pieces of evidence about my financial position, about my other buying opportunities, about my apparent bargaining strategy. In "House Sale," unlike "Quaker House Sale," you do not trust me. You retain the full independence of your judgment, and do not give me sufficient control over matters so that I might jeopardize your autonomous choice. Because your autonomy need not be at risk and because I have a genuine need to protect my position, I may deceive you without wronging you, it seems.

I do not expect everybody to agree with me about the possibility of acceptable deception in cases like "House Sale." I am instead concerned with how trust matters in assessing deception, given that deception can occur even in the absence of trust. My suggestion on this score is simple. It is always morally unacceptable to deceive a person in a way that breaches his trust, unless that deception is necessary to defend against a grave wrong. But it may be morally acceptable to defend a person in the absence of trust if that deception is necessary to defend against an action that may thwart one's legitimate interests. One might reject my suggestion and still see an important moral difference between deception involving a breach of trust and simple deception. One might thus say that other things being equal, deception involving a breach of trust is morally worse than deception by other manipulative means. No doubt deception by breach of trust is commonly worse that deception by more simple manipulation. But that contrast does not track the complexity of the difference between these two kinds of deception. Indeed, there are cases in which both kinds of deception would be morally acceptable, neither wrong, and hence neither worse than the other; for example, the case in which one seeks to fend off the murderer at the door making queries about the whereabouts of his prospective victim. The morally important difference between deception that occurs through breach of trust and deception that occurs through more simple manipulation, I have argued, is that there are different rationales appropriate for employing each kind of deception.

ALAN STRUDLER

University of Pennsylvania

COMMENTS AND CRITICISM

NEW DISTINCTIONS, SAME TROUBLES: A REPLY TO HAJI AND MCKENNA*

Ishtiyaque Haji and Michael McKenna[1] present an argument that, contrary to the received view, a prior-sign Frankfurt case that assumes causal determinism does not beg the question against the incompatibilist and can support the conclusion that moral responsibility does not require alternative possibilities.[2] In this paper, I aim to show that Haji and McKenna's argument fails and a prior-sign Frankfurt case that assumes determinism does beg the question against the incompatibilist.

I. PRIOR-SIGN FRANKFURT CASES

The kind of Frankfurt case Haji and McKenna consider is the prior-sign type.[3] Here is their example case.

> Mork believes that in his circumstances it is wrong for him to lie to prevent embarrassment. In spite of this, he decides to lie and lies anyway. As he has no excuse or justification for lying, it seems he is deserving of blame for lying. Unbeknowst to Mork, though, he could not have refrained from lying due to the presence of Mindy who has the power to read and control Mork's mind. Mindy wields this power partly in virtue of possessing the following knowledge. Had Mork been about to refrain from lying, he would have displayed some involuntary sign—a neurological pattern, *N**, in his brain—whereas if he had been about to lie, he would have displayed a different neurological pattern, *N*. Had Mindy detected *N**, she would have interceded in Mork's deliberations via direct stimulation of Mork's brain and would, in this way, have caused Mork to lie. But Mindy detects *N*, the reliable sign for Mindy that she need not show her hand at all. As Mork lies on his own, in the absence of Mindy's intervention, it seems highly reasonable that Mork acts freely and is morally blameworthy for lying despite not having alternative possibilities (300).

---

*I should like to thank Robert Kane for comments on an earlier draft and helpful discussion.

[1] "Dialectical Delicacies in the Debate about Freedom and Alternative Possibilities," this JOURNAL, CI, 6 (June 2004): 299–314.

[2] I will switch in the rest of the paper between "causal determinism" and "determinism." This is just for exegetical convenience. I treat the terms synonymously.

[3] From now on I will drop the "prior-sign" locution and just refer to "Frankfurt cases," although these are the only kinds I am concerned with here.

0022-362X/05/0209/474–82

This situation seems a problem for those wanting to hold that moral responsibility requires that the agent have alternative possibilities. Many philosophers, particularly incompatibilists, have tried to argue that—contrary to the intuition—the Frankfurt cases do not show that one can be morally responsible without having any alternative possibilities. A popular strategy to do this, first developed by Robert Kane[4] and David Widerker,[5] is to claim that careful consideration of the cases points to a dilemma, both horns of which are problematic for those wanting to defend Frankfurt's conclusions. Following Haji and McKenna, I will call those who believe the Frankfurt cases establish that alternative possibilities are not necessary for moral responsibility, *Frankfurt Defenders*; and I will refer to those who claim the dilemma forces the rejection of Frankfurt's conclusions, *Dilemma Defenders*. I will now present a standard formulation of the dilemma defense.

II. THE DILEMMA DEFENSE

A standard way of presenting the dilemma is to consider what follows if the Frankfurt case assumes either causal determinism or indeterminism. Assuming causal determinism obtains in the Frankfurt cases is to assume that there is a determined relationship between the prior sign Mork exhibits and his choice to act. This assumption—the first horn of the dilemma—begs the question against the incompatibilist who does not think that causal determinism is compatible with free will and moral responsibility, and so, from the outset, will not be able to hold Mork morally responsible for his action. The other horn of the dilemma covers the assumption of causal indeterminism obtaining in the Frankfurt case. While Mork might be held morally responsible for his action under such conditions, there is no longer a causally determined relationship between the prior sign and his choice since causal indeterminism denies this. Therefore, the only way in which Mindy could intervene would be after Mork has already begun to lie—surely a robust alternative possibility. If she intervened before that, Mindy would be determining the outcome and would be responsible. She would be an actual, and not mere counterfactual, intervener.

Much of the response to this dilemma in the literature has focused on the indeterminism horn. Indeed, the main compatibilist project has been to construct cases in which an agent is morally responsible for his action though he has no alternative possibilities and acts in an indeterministic setting (see Alfred Mele and David Robb[6] for examples of such cases). By contrast, little attention has been paid until recently to the determinism horn. That is, Frankfurt Defenders have tended not to continue to construct prior-sign Frankfurt cases within a causally determined setting. An obvious reason for this is common acceptance that these sorts of cases really *are* question-begging against the incompatibilist.

---

[4] *Free Will and Values* (Albany: SUNY, 1985).

[5] "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *Philosophical Review*, CIV (1995): 247–61.

Haji and McKenna's goal, however, is to challenge the question-begging charge. More specifically, they attempt to analyze the causal determinism assumption carefully to see what it entails. They claim it is possible for the Frankfurt Defender to overcome the question-begging charge. That is, Haji and McKenna claim that causal determinism can be assumed within a Frankfurt case without begging the question against the incompatibilist. I want to show their argument for this fails.

III. HAJI AND MCKENNA'S DISTINCTION AND THE "BROAD" DIALECTICAL CONTEXT

Recall that the question-begging charge goes like this. If a causally determined relationship is assumed between the prior sign and Mork's choice to act, then the incompatibilist cannot hold Mork morally responsible for his action, since the incompatibilist does not believe determinism is compatible with free will and moral responsibility. In a situation designed to test requirements for free will and moral responsibility, this assumption begs the question against the incompatibilist. Haji and McKenna begin by distinguishing two readings of the question-begging charge: a broad, and narrow, interpretation. The *broad* interpretation, they claim, "begs the question against the incompatibilist *because a deterministic relation obtains—simpliciter*" (303). This is to say that determinism is incompatible with free will and moral responsibility for the simple reason that determinism itself is incompatible. As a contrast, they argue the *narrow* interpretation "begs the question against the incompatibilist *because the deterministic relation expunges alternative possibilities*" (303). Here the claim is not that the situation is question-begging because determinism itself is incompatible with free will and moral responsibility. Rather, it is the absence of alternative possibilities, which determinism entails, that is incompatible with free will and moral responsibility.

Each of the two interpretations, Haji and McKenna suggest, have different audiences to which they are appropriately aimed. The broad reading concerning determinism itself is "most reasonably suited for a dialectic in which Frankfurt's argument is directed at a philosophically *committed audience*" (302). By this, they mean an audience who firmly believes determinism itself is incompatible with free will and moral responsibility. On the other hand, the narrow interpretation focusing on the role of alternative possibilities in moral responsibility is "most reasonably suited for a dialectic in which Frankfurt's argument is directed at an *undecided audience*" (302). Haji and McKenna claim that if the incompatibilist is to be fair to the Frankfurt Defender, he must suspend judgment on whether alternative possibilities are required for free will and moral responsibility. Hence, they say the narrow interpretation ought to be aimed at an undecided audience. Haji and McKenna feel this distinction is important as it both gives some insight into the kind of response Frankfurt Defenders might give, and it outlines the commitments of the Dilemma Defender. They draw two main conclusions.

---

[6] "Rescuing Frankfurt-style Cases," *Philosophical Review*, CVII (1998): 97–112.

First, Haji and McKenna admit that, under the broad reading, "the Dilemma Defender is entitled to the complaint that the first horn of the dilemma [that is, assuming determinism] is question-begging [against incompatibilists]" (313). This is because the broad interpretation objects that determinism *simpliciter* obtains in the example. If the Frankfurt Defender wants to work under the broad interpretation, he must offer an example that does not assume the agent is determined (as, for instance, Mele and Robb do). Haji and McKenna claim that if an incompatibilist is willing to consider an indeterministic Frankfurt case, he must suspend judgment on the role of alternative possibilities. The incompatibilist, in such an instance, is "not entitled to demand that (absent Frankfurt's counterfactual machinery) the relevant case involves an agent with alternative possibilities. Demanding this is tantamount to refusing Frankfurt a fair hearing" (311).

The upshot—under the rules of considering an indeterministic Frankfurt example—is that Dilemma Defenders are "dialectically committed at least to the possibility that determinism might rule out free will and moral responsibility for reasons other than ruling out alternative possibilities" (313). Furthermore, Haji and McKenna argue that "if incompatibilists wish to grapple with Frankfurt given a broad dialectical situation, we believe they do so in bad faith. By demanding that Frankfurt begin with a case of a libertarian free agent, they are demanding that Frankfurt provide a case in which any freely willing and morally responsible agent does have alternative possibilities. This of course is just what Frankfurt wants to disprove" (313).

Given what Haji and McKenna have said about the incompatibilist's dialectical commitments and his bad faith in pursuing the broad dialectical situation, they claim incompatibilists must concede "they are not entitled to the dilemma defense if it is meant to address a broad dialectical situation. If these incompatibilists wish to rely upon the dilemma defense, then they must accept the conditions of the debate established by the narrow dialectical context" (312).

IV. THE "NARROW" DIALECTICAL CONTEXT

Haji and McKenna's discussion of the broad dialectic yields two main conclusions. First, Frankfurt Defenders must grant the question-begging nature of assuming determinism, and try to prove their point only by using Frankfurt cases that assume indeterminism. Second, incompatibilists should consider cases in the narrow context, and so give Frankfurt a fairer hearing. The point of the second conclusion seems to me to be that if the Dilemma Defender is forced to consider Frankfurt cases under the narrow context, the dilemma defense cannot work. Haji and McKenna believe the determinism assumption is not question-begging (the Dilemma Defender's charge) and so it seems that the Dilemma Defender's claim falls away, entailing Frankfurt's original conclusions have to be accepted. However, I do not believe their argument works. They do not show that determinism can be assumed in a nonquestion-begging manner.

Recall, the narrow dialectic is aimed at a philosophically uncommitted audience, and the question-begging charge is aimed specifically at how the assumption of determinism rules out alternative possibilities before the counterfactual intervener is introduced. The idea is that an uncommitted audience might think the Frankfurt Defender is begging the question against the incompatibilist not simply by assuming determinism *simpliciter*, but by the fact that alternative possibilities are eliminated by determinism even prior to the introduction of the counterfactual intervener. The counterfactual intervener is doing no work in the case since determinism already rules out alternative possibilities. A common motivation for incompatibilism is the feeling that in a determined environment, given the prior causal history and laws of nature, at the moment of choice, an agent cannot truly possess alternative possibilities. And if an agent does not have alternative possibilities, how can the choice be *his*? As soon as determinism is assumed, the notion of the agent having alternative possibilities is ruled out. Haji and McKenna present the Dilemma Defenders' claims (under the guise of the uncommitted philosopher), according to the narrow reading, as being "to insure an important question is not begged...we need a case which the *only* manner of ruling out alternative possibilities is by way of the counterfactual intervener" (309).

Haji and McKenna agree that "the Frankfurt Defender's case would be more convincing—there would be less to contest—if it were beyond dispute that an agent's alternatives were ruled out *only* by virtue of the counterfactual intervener and not by virtue of the truth of determinism" (310). However, they do not believe it is correct to say the Frankfurt Defender is begging the question in his argument. Haji and McKenna claim the Frankfurt Defender has a nonquestion-begging reply. I will present their reply in three stages so as to get clear about what is being said.[7]

(1) The counterfactual intervener and causal determinism are individually sufficient to rule out alternative possibilities.

"The counterfactual intervener's presence in a deterministic scenario, while perhaps not necessary for ruling out alternative possibilities, is sufficient for doing so" (310). In other words, "even if determinism rules out alternative possibilities, so does the presence of Frankfurt's counterfactual machinery" (313).

(2) The counterfactual intervener's ruling out of alternative possibilities is irrelevant in judging the agent not morally responsible.

"Furthermore, [the counterfactual intervener's ruling out of alternative possibilities] is irrelevant to any judgment that the agent is not free or morally responsible" (310).

(3) Determinism's ruling out of alternative possibilities is irrelevant as well for judging the agent is not morally responsible.

"The [counterfactual intervener] way of ruling out alternatives is irrelevant to an evaluation of an agent's freedom and responsibility, and this provides plausible grounds for the suggestion that [the assumption of determinism] is as well" (313).

To summarize, Haji and McKenna's reply on behalf of the Frankfurt Defender seems to be:

(1) Determinism and the counterfactual intervener are individually sufficient for ruling out alternative possibilities.
(2) The counterfactual intervener way of ruling out alternative possibilities is irrelevant to judging whether an agent is free and has moral responsibility.
(3) So, from (1) and (2), the determinism way of ruling out alternative

---

[7] To complicate matters somewhat, Haji and McKenna present the reply on two occasions in their paper (310 and 313) but the later presentation is the fuller account of the reply. In many ways, though, the earlier rendering is the easiest to follow. This accounts for the chopped up quotations in my presentation of their idea.

> possibilities is irrelevant to judging whether an agent is free and has moral responsibility.

This conclusion, say Haji and McKenna, is not question-begging. Recall the narrow dialectic charge of question-begging focused on the deterministic relation ruling out alternative possibilities independently of the counterfactual intervener.

However, this argument is not sound. (1) is false. Determinism and the counterfactual intervener are *not* individually sufficient for ruling out alternative possibilities. Only determinism is sufficient for ruling out alternative possibilities. The counterfactual intervener can only rule out alternative possibilities in conjunction with an assumption that determinism holds. Put another way, the counterfactual intervener *qua* counterfactual intervener is not sufficient to rule out alternative possibilities. It only rules out alternative possibilities in a determined environment. To see that this is so, let us return to my presentation of the dilemma at the beginning of the paper. For the counterfactual intervener (Mindy, in the example) to pick up on the "prior sign" as a reliable indicator of the agent's (Mork's) choice and action, determinism must hold, since indeterminism would deny a reliable relationship between the two. Imagine Mork decides not to lie. In order for Mindy to intervene successfully, without Mork's knowledge, determinism must hold, since Mindy detects the prior sign and activates different neural pathways sufficient for Mork actually to lie. However, if we assume an undetermined environment, there will be no determined relationship between the prior sign and the choice. So, if Mindy is to intervene, she will have to wait until Mork has already, in some sense, chosen to lie before being able to alter Mork's neural mechanisms. The Frankfurt Defender wants, of course, to rule out such a possibility for this is an instance of a robust alternative possibility. So, for the counterfactual intervener to rule out alternative possibilities, which would involve him picking up on the prior sign, the relevant Frankfurt case must assume determinism holds. To repeat, the counterfactual intervener can only rule out alternative possibilities in an environment that assumes determinism. Because of this, it is incorrect to claim that the counterfactual intervener alone is sufficient for expunging alternative possibilities. Only determinism is sufficient for this. In other words, the counterfactual intervener together with determinism is sufficient but that is because determinism is already sufficient in itself.

I now want to say more about why the falsity of (1) undermines Haji and McKenna's argument. The truth of (3), which effectively purges the Frankfurt Defender of the question-begging charge, rests on the parallel presented in (1)—that is, the claim that determinism and the counterfactual intervener are individually sufficient to rule out alternative possibilities. If (1) is not true, then (3) does not follow, and the question-begging charge will not have been overcome. If we accept (1), and believe the counterfactual intervener is irrelevant to judging the moral responsibility of the agent, it follows that determinism would also be irrelevant to judging the agent not morally responsible. The way this conclusion about determinism is established does not seem to beg any questions against the incompatibilist. However, given that (1) is false, (3) does not follow. In other words, it is not the case, contrary to what (3) asserts, that the determinism way of ruling out alternative possibilities is irrelevant to judging whether agent is free and has moral responsibility. In fact, the way determinism rules out alternative possibilities, prior to the introduction of a counterfactual intervener, is going to be relevant to judging if the agent is free and morally responsible for his choice and action. Recall, the narrow dialectic begs the question against the incompatibilist because the deterministic relation expunges alternative possibilities. Haji and McKenna try to show that determinism can be assumed without begging any questions. However, what I have shown is that the way in which Haji and McKenna try to avoid the question-begging charge does not work. I still maintain therefore, that even under the narrow interpretation, the assumption of determinism begs the question against the incompatibilist. The only way, it seems to me, to avoid the question-begging is for the Frankfurt defender to go back to the broad dialectic and construct a case that assumes indeterminism. (This is, of course, the technique Mele and Robb and others employ.)

V. CONCLUSION

I have tried to show that Haji and McKenna are mistaken in claiming that the Frankfurt Defender can assume determinism holds in a Frankfurt case in a nonquestion-begging manner. There is no way, it seems to me, to construct a prior-sign Frankfurt case that assumes determinism and does not beg the question against the incompatibilist. What the Frankfurt defender must do, therefore, is move to challenge incompatibilists on the other horn of the dilemma. They must construct cases that assume indeterminism obtains. This recommendation is in line with that of Haji and McKenna's conclusions after their assessment of the broad dialectical reading. They made this recommendation while believing that Frankfurt Defenders could still appeal to the narrow context to save the assumption of determinism. However, now that I have shown their strategy fails, I challenge Frankfurt Defenders to adopt Haji and McKenna's insistence that if they want

BOOK REVIEWS

Truth and Paradox: Solving the Riddles. TIM MAUDLIN. New York: Oxford University Press, 2004. xii + 209 p. Cloth $74.00.

Attempts to give a consistent account of the liar paradox are not exactly new. Logicians have been attempting them for well over two millennia, but not with a great deal of success—at least if consensus is any reasonable guide. This fact does not exactly encourage optimism about new attempts; but of course there may well be some hitherto untried way of doing the job. In this book, Tim Maudlin turns his hand (from the philosophy of physics) to the challenge.

The book is clearly written and argued, and certainly contains interesting insights. It is also a very honest book: difficulties for the view espoused are not pushed under the carpet, but faced squarely. Some bits may strike a logician as a little odd. For example, Maudlin says, "It is, I am told, something like folk wisdom among logicians that the [arguments in the liar paradox] are to be defeated by disallowing the *T*-inferences [α + 'α' is true] in conditional proofs, i.e., within a subderivation" (106, and again, 132). I think that this will come as a surprise to most of the folk. He also says, "It is not clear how...Relevance Logic or Paraconsistent Logic could block any of [the inferences involved in Curry's paradox]" (15). It is a pity he did not seek advice on this as well: contraction-free relevant and paraconsistent logics are not exactly esoterica. On the other hand, the book contains nice illustrations of various ideas drawn from physics that would probably not have occurred to a logician.

In chapter 1, Maudlin spells out the liar paradox and Löb's paradox (or Curry's paradox, as it is perhaps more usually called). Chapters 2 and 3 spell out the views about truth on which Maudlin's solution to the paradoxes depends. Chapter 4 presents a language that is claimed to be "semantically closed." For reasons we will come to in due course, the account of truth needs to be augmented by a theory of permissible assertion; this is given in chapter 5. Chapter 6 gives a proof-theory suitable for the notion of truth involved, and applies it to some of the arguments related to the liar paradox thrown up by Gödel's Theorem (which is the topic, Maudlin tells us in the preface to the book, which generated his interest in the area). Chapter 7 gives a proof-theory for permissible assertability, which allows him to discuss an obvious extended paradox in chapter 8. Finally, chapter 9 replies to another very natural objection to the whole approach.

0022-362X/05/0209/483–86

Maudlin's solution comes in several distinct phases. The first is the claim that the liar sentence is neither true nor false. To justify this, Maudlin endorses the process by which sentences receive truth values in the construction of Saul Kripke's minimal fixed point (with Kleene 3-valued logic), so validating the "*T*-inferences." He then adds the claim that this is the only process by which truth values may be attained. So anything that is truth-valueless at the minimal fixed point is truth-valueless, *period.*

Whether this solves the liar paradox, we will come back to in a minute. The strategy would appear to be in trouble with some closely related paradoxes (not discussed by Maudlin), which one would expect to have the same sort of solution as the liar. Consider the "knower paradox"—κ: 'κ' is not known (by me). If this satisfies Excluded Middle, we have the usual contradiction. Now, belief is a nonalethic matter, and so statements about it are already determined as either true or false in the ground model of the construction—Maudlin calls such sentences "boundary sentences." Suppose that I do not, as a matter of fact, believe κ (which, in fact, I do not). Assuming that knowledge is some kind of true belief, it follows that I do not know it; so it *is* true; *a fortiori,* it satisfies Excluded Middle. (And if knowledge is *sui generis,* then it is unclear how one might go about applying Maudlin's strategy to solve it. Nor, for that matter, is it clear how the strategy might be applied to semantic paradoxes such as Berry's, which do not use the Law of Excluded Middle.)

Returning to the liar paradox itself, there is an obvious objection to the solution. According to the account in question, the liar sentence is not true, but that very claim is not true. So the theory would seem to be self-refuting. Nearly all the things one would want to say about truth by way of a theory of truth are not true either, such as that every conjunction is true if both conjuncts are true—when the quantifier is instantiated with a truth-valueless sentence, this is neither true nor false—or even that nothing true is false. In chapter 4, Maudlin constructs, in a natural way, a truth theory for a language with a truth predicate, and himself points out that, by his own lights, the theory is not true. (Actually, there are some technical problems with Maudlin's theory. The truth theory is for a first order language, but is given in a second order language with quantifiers over functions, so it is not semantically closed.)

In an attempt to solve this problem we come to the second phase of Maudlin's strategy. We invoke a notion of what it is permissible to assert. The norms involving this notion are, *inter alia,* as follows (97–98):

- all truths may be asserted;
- no falsehoods may be asserted;

- an ungrounded atomic sentence may not be asserted;
- the negation of an ungrounded atomic sentence may be asserted.

Thus, the liar sentence—λ: 'λ' is not true—may be asserted, but the claim that it is true may not.

An obvious question at this point is what to make of the notion of assertion and its norms. Maudlin seems to think (95) that standard norms for assertion are simply silent on what to do with truth-valueless sentences, and that we are therefore at liberty to fill in the silences as suits us best. But this is far from the case. The natural norm for assertability is that α may be asserted if it is true, and may not be asserted if it is not true. Truth is the aim of assertion. Once this connection is broken, the notion of assertion comes free from its mooring, and it is not clear why we should assert anything. Okay: we can assert that the liar is not true, but so what? Since it is not true, there seems to be no reason why I should believe it.

And just because this connection is broken, the norms concerning truth-valueless sentences proposed by Maudlin appear to be arbitrary. Why not assert atomic ungrounded sentences, but not their negations? Because, of course, that gives intuitively the wrong answer. But we are now bereft of an explanation as to why it is wrong.

Worse is to come. The introduction of this new machinery into the equation allows us, as usual, to formulate other self-referential paradoxes. The obvious one now is the sentence π: it is not permissible to assert 'π'. This is either true or false, since it is a boundary sentence. But if it is true it cannot be asserted, and if it is false, it can be asserted. One way or another, the rules for assertion have become incoherent. The connection between assertability and truth has broken, even in the case of sentences that are truth valued. Maudlin's response (chapter 8) is to shrug his shoulders. We know that following rules can sometimes get us into trouble. This is just a case of that. Phase three of the strategy.

But even here things are not that simple. If something is true, it is permissible to assert it; if something is untrue, it is not permissible to assert it. While we maintain these claims, π still generates a contradiction: it is and is not permissible to assert π. Of course, we may give up these claims. (Mauldin, in fact, "baldly denies" them (189).) But how is this any better than giving up the natural principles of inference concerning truth? A possible reply is to the effect that the *T*-inferences are constitutive of the notion of truth. If it is *truth* that we are talking about, one cannot give them up. But, for all the world, the claims about assertability seem to be constitutive of the notion of assertability. If it is *assertability* that we are talking about, one cannot give them up.

There is one final bullet to be bitten. The liar sentence is a contradiction: it is not possible for it to be true. It is none the less assertable. So, as Maudlin himself points out in a different context (127), his account commits us to the assertion of contradictory sentences. In the end, then, we have to accept a contradiction. The contradiction may not be true, but this is somewhat cold comfort. Assertability was introduced as a surrogate for truth, precisely because, on Maudlin's account, some of the things he wants to endorse are not true. It is a way of hanging on to the Good Things. It now turns out that contradictions can be Good Things too.

Maudlin's journey in search of the Holy Grail of avoiding contradiction has been a long one. On the way, there have been many casualties, *ad hoc* maneuvers, and hostages given up. And in the end, contradiction has not been avoided, only its truth—but the truth about semantics turned out to be rather uninteresting anyway: most things one might expect a theory of truth to deliver are untrue. It was a gallant try, and well worth the effort of exploring the approach. But I think that the main thing that Maudlin has taught us—though I am sure that he does not agree with this (and thanks to him for a generous set of comments on a first draft of the review)—is that yet another strategy for being consistent about the liar paradox fizzles out.

GRAHAM PRIEST

Universities of Melbourne and St Andrews

NEW BOOKS

ALEXANDER, LARRY. *Is There a Right of Freedom of Expression?* New York: Cambridge, 2005. xii, 203 p. Cloth $70.00, paper $24.99.

ALLEN, ABRAM: *Truth Eternal and the Adversity of Diversity Law.* Lanham, MD: Hamilton, 2005. xxii, 323 p. Paper $40.00.

ALLIEZ, ERIC: *The Signature of the World.* New York: Continuum, 2004. xxix, 120 p. Cloth $105.00, paper $19.95.

ALSTON, WILLIAM P.: *Beyond "Justification": Dimensions of Epistemic Evaluation.* Ithaca, NY: Cornell, 2005. xv, 256 p. Cloth $45.00.

AMUNDSON, RON: *The Changing Role of the Embryo in Evolutionary Thought.* New York: Cambridge, 2005. xiii, 280 p. Cloth $75.00.

APPLEBAUM, WILBUR: *The Scientific Revolution and the Foundations of Modern Science.* Westport, CT: Greenwood, 2005. xxi, 243 p. Cloth $45.00.

ARIETI, JAMES A.: *Philosophy in the Ancient World: An Introduction.* Lanham, MD: Rowman and Littlefield, 2005. xxv, 386 p. Cloth $80.00, paper $28.95.

ATLAS, JAY DAVID: *Logic, Meaning, and Conversation: Semantical Underdeterminacy, Implicature, and Their Interface.* New York: Oxford, 2005. xv, 284 p. Cloth $65.00.

AUDI, ROBERT: *The Good in the Right: A Theory of Intuition and Intrinsic Value.* Princeton: University Press, 2004. xi, 244 p. Cloth $45.00, paper $22.95.

BAKER, ROBERT: *The Extravagant: Crossings of Modern Poetry and Modern Philosophy.* Notre Dame: University Press, 2005. xvii, 405 p. Cloth $65.00, paper $30.00.

BALSWICK, JACK O., PAMELA EBSTYNE KING, and KEVINS REIMER: *The Reciprocating Self: Human Development in Theological Perspective.* Downers Grove, IL: InterVarsity Press, 2005. 334 p. Paper $24.00.

BARBER, STEPHEN: *Jean Genet.* London: Reaktion, 2004. 155 p. Paper $16.95.

BAR-ON, DORIT: *Speaking My Mind: Expression and Self-Knowledge.* New York: Oxford, 2004. xiii, 449 p. Paper $45.00.

BATTIN, MARGARET PABST: *Ending Life: Ethics and the Way We Die.* New York: Oxford, 2005. 344 p. Cloth $74.00, paper $24.95.

BELSHAW, CHRISTOPHER: *Ten Good Questions about Life and Death.* Malden, MA: Blackwell, 2005. vii, 178 p. Paper $19.95.

BENHABIB, SEYLA: *The Rights of Others: Aliens, Residents, and Citizens.* New York: Cambridge, 2004. xii, 251 p. Cloth $65.00, paper $23.99.

BERG, MAXINE: *Luxury and Pleasure in Eighteenth-Century Britain.* New York: Oxford, 2005. xvii, 373 p. Cloth $45.00.

BINMORE, KEN: *Natural Justice.* New York: Oxford, 2005. xiii, 207 p. Cloth $29.95.

BISHOP, MICHAEL A., and J.D. TROUT: *Epistemology and the Psychology of Human Judgment.* New York, Oxford, 2004. Cloth $74.00, paper $24.95.

BORG, EMMA: *Minimal Semantics.* New York: Oxford, 2004. x, 288 p. Cloth $74.00.

BOWIE, NORMAN E., with PATRICIA H. WERHANE: *Management Ethics.* Malden, MA: Blackwell, 2005. viii, 159 p. Cloth $61.95, paper $26.95.

BOYD, ROBERT, and PETER J. RICHERSON: *The Origin and Evolution of Cultures.* New York: Oxford, 2005. viii, 456 p. Cloth $75.00, paper $35.00.

BRENNAN, TAD: *The Stoic Life: Emotions, Duties and Fate.* New York: Oxford, 2005. xi, 340 p. Cloth $45.00.

BROOME, JOHN: *Weighing Lives.* New York: Oxford, 2004. ix, 278 p. Cloth $35.00.

BROWN, CHRISTOPHER M.: *Aquinas and the Ship of Theseus: Solving Puzzles about Material Objects.* New York: Continuum, 2005. ix, 194 p. Cloth $120.00.

BULLER, DAVID J.: *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature.* Cambridge: MIT, 2005. xi, 550 p. Cloth $34.95.

BURGE, TYLER: *Truth, Thought, Reason: Essays on Frege.* New York: Oxford, 2005. xii, 419 p. Cloth $110.00, paper $24.95.

BURGESS, JOHN P.: *Fixing Frege.* Princeton: University Press, 2005. 257 p. Cloth $39.95.

BURIAN, RICHARD: *The Epistemology of Development, Evolution, and Genetics: Selected Essays.* New York: Cambridge, 2005. xiii, 274 p. Cloth $75.00, paper $32.99.

BURNHAM, DOUGLAS: *Kant's Philosophies of Judgement.* Edinburgh: University Press, 2004. vii, 213 p. Cloth $90.00.

BUTLER, REX: *Slavoj Žižek: Live Theory.* New York: Continuum, 2005. viii, 165 p. Cloth $89.95, paper $19.95.

BYRNE, RUTH M.J.: *The Rational Imagination: How People Create Alternatives to Reality.* Cambridge: MIT, 2005. xiv, 254 p. Cloth $32.00.

CAHN, STEVEN M.: *God, Reason, and Religion.* Belmont, CA: Thomson Wadsworth, 2006. xii, 94 p. Paper $17.95.

CAHOONE, LAWRENCE E.: *Cultural Revolutions: Reason versus Culture in Philosophy, Politics, and Jihad.* University Park: Penn State UP, 2005. 231 p. Cloth $35.00.

CAMPBELL, NEIL: *A Brief Introduction to the Philosophy of Mind.* Orchard Park, NY: Broadview, 2005. 213 p. Paper $22.95.

CAPPELEN, HERMAN, and ERNIE LEPORE: *Insensitive Semantics.* Malden, MA: Blackwell, 2005. xii, 219 p. Cloth $64.95, paper $29.95.

CARTER, STEVEN: *The Nothing That Is and the Nothing That Is Not: On Death, Dying, and Suffering.* Lanham, MD: University Press of America, 2004. xix, 128 p. Paper $24.00.

CEREZO, MARIA: *The Possibility of Language: Internal Tensions in Wittgenstein's Tractatus.* Chicago: University Press, 2005. xiv, 321 p. Paper $30.00.

CHAMBERS, SAMUEL A.: *Untimely Politics.* New York: University Press, 2003. xi, 196 p. Cloth $45.00.

CHANG, HASOK: *Inventing Temperature: Measurement and Scientific Progress.* New York: Oxford, 2004. xviii, 286 p. Cloth $49.95.

CHAPMAN, SIOBHAN: *Paul Grice: Philosopher and Linguist.* New York: Palgrave Macmillan, 2005. viii, 247 p. Cloth $74.95.

CHAPPELL, TIMOTHY: *Reading Plato's Theaetetus.* Indianapolis: Hackett, 2005. 246 p. Paper $16.95.

CHRISTENSEN, DAVID: *Putting Logic in Its Place: Formal Constraints on Rational Belief.* New York: Oxford, 2004. xii, 187 p. Cloth $49.95.

CLIFTON, ROB: *Quantum Entanglements: Selected Papers.* JEREMY BUTTERFIELD and HANS HALVORSON, eds. New York: Oxford, 2004. xliii, 461 p. Cloth $120.00.

CLOUSER, ROY A.: *The Myth of Religious Neutrality, Revised Edition.* Notre Dame: University Press, 2005. xii, 397 p. Paper $25.00.

COHEN, STEPHEN: *The Nature of Moral Reasoning.* New York: Oxford, 2004. ix, 163 p. Paper $24.95.

CONEE, EARL, and THEODORE SIDER: *Riddles of Existence: A Guided Tour of Metaphysics.* New York: Oxford, 2005. vii, 210 p. Cloth $17.95.

CORAZZA, EROS: *Reflecting the Mind: Indexicality and Quasi-Indexicality.* New York: Oxford, 2004. ix, 368 p. Cloth $65.00.

CORRIGAN, KEVIN: *Reading Plotinus.* West Lafayette, IN: Purdue, 2004. xiii, 293 p. Cloth $58.95, paper $28.95.

CRITCHLEY, SIMON: *Things Merely Are.* New York: Routledge, 2005. viii, 137 p. Cloth $80.00, paper $22.95.

CRIVELLI, PAOLO: *Aristotle on Truth.* New York: Cambridge, 2004. xi, 340 p. Cloth $85.00.

CULLITY, GARRETT: *The Moral Demands of Affluence.* New York: Oxford, 2004. ix, 286 p. Cloth $60.00.

CURRIE, GREGORY: *Arts and Minds.* New York: Oxford, 2005. x, 281 p. Cloth $99.00, paper $35.00.

CUTROFELLO, ANDREW: *Continental Philosophy: A Contemporary Introduction.* New York: Routledge, 2005. xvi, 440 p. Paper $29.00.

DANCY, JONATHAN: *Ethics Without Principles.* New York: Oxford, 2004. x, 224 p. Cloth $45.00.

DAVIDSON, DONALD: *Truth, Language, and History.* New York: Oxford, 2005. xx, 350 p. Cloth $74.00, paper $24.95; *Truth and Prediction.* Cambridge: Harvard, 2005. x, 180 p. Cloth $29.95.

DEARDEN, IAN: *Do Philosophers Talk Nonsense? An Inquiry into the Possibility of Illusions of Meaning.* England: Bank House Books, 2005. 136 p. Paper $22.99.

DE BEAUVOIR, SIMONE: *Philosophical Writings.* MARGARET A. SIMONS, ed. Urbana: Illinois UP, 2004. xii, 351 p. Cloth $40.00.

DEGRAZIA, DAVID: *Human Identity and Bioethics.* New York: Cambridge, 2005. xi, 300 p. Cloth $70.00, paper $25.99.

DEN HARTOGH, GOVERT: *Mutual Expectations: A Conventionalist Theory of Law.* New York: Kluwer, 2002. xi, 288 p. Cloth $122.00.

DENNETT, DANIEL C.: *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness.* Cambridge: MIT, 2005. xiii, 199 p. Cloth $28.00.

DERSHOWITZ, ALAN: *Rights from Wrongs.* New York: Basic, 2004. ix, 261 p. Cloth $24.00.

DE VRIES, HENT: *Minimal Theologies: Critiques of Secular Reason in Adorno and Levinas.* Baltimore: Johns Hopkins, 2005. xxxiv, 720 p. Cloth $65.00, paper $24.95.

DICKER, GEORGES: *Kant's Theory of Knowledge.* New York: Oxford, 2004. xv, 262 p. Cloth $74.00, paper $19.95.

DIETRICH, ERIC, and VALERIE GRAY HARDCASTLE: *Sisyphus's Boulder: Consciousness and the Limits of the Knowable.* Philadelphia: John Benjamins, 2004. xii, 133 p. $108.00.

DI GIOVANNI, GEORGE: *Freedom and Religion in Kant and His Immediate Successors.* New York: Cambridge, 2005. xvi, 373 p. Cloth $75.00.

DISTIN, KATE: *The Selfish Meme.* New York: Cambridge, 2005. v, 231 p. Cloth $60.00, paper $21.99.

DOWE, PHIL: *Galileo, Darwin, and Hawking: The Interplay of Science, Reason, and Religion.* Grand Rapids, MI: Eerdmans, 2005. viii, 205 p. Paper $21.00.

DUNCAN, CRAIG, and TIBOR R. MACHAN: *Libertarianism: For and Against.* Lanham, MD: Rowan and Littlefield, 2005. xix, 169 p. Cloth $49.95, paper $16.95.

DUSTIN, CHRISTOPHER A. and JOANNA E. ZIEGLER: *Practicing Mortality: Art, Philosophy, and Contemplative Seeing.* New York: Palgrave Macmillan, 2005. xiv, 253 p. Cloth $65.00.

ELLIS, ELIZABETH: *Kant's Politics: Provisional Theory for an Uncertain World.* New Haven: Yale, 2005. xii, 260 p. Cloth $40.00.

ELLIS, RALPH: *Love and the Abyss: An Essay on Finitude and Value.* Chicago: Open Court, 2004. viii, 271 p. Paper $25.95.

EMDEN, CHRISTIAN J.: *Nietzsche on Language, Consciousness, and the Body.* Urbana: Illinois UP, 2005. xii, 223 p. Cloth $35.00.

ETSIONI, AMITAI: *The Common Good.* Malden, MA: Polity, 2004. vii, 251 p. Cloth $64.95, paper $27.95.

EVANS, J. CLAUDE: *With Respect for Nature.* Albany: SUNY, 2005. xxi, 280 p. Cloth $73.50, paper $21.95.

EVENSKY, JERRY: *Adam Smith's Moral Philosophy: A Historical and Contemporary Perspective on Markets, Law, Ethics, and Culture.* New York: Cambridge, 2005. xv, 331 p. Cloth $75.00.

FETZER, JAMES H: *The Evolution of Intelligence: Are Humans the Only Animals with Minds?* Chicago: Carus, 2005. xx, 272 p. Paper $34.95.

FIALA, ANDREW: *Tolerance and the Ethical Life.* New York: Continuum, 2005. 194 p. Cloth $105.00.

FITZPATRICK, JOSEPH: *Philosophical Encounters: Lonergan and the Analytic Tradition.* Toronto: University Press, 2005. vii, 233 p. Cloth $65.00, paper $29.95.

FLATHMAN, RICHARD E.: *Pluralism and Liberal Democracy.* Baltimore: John Hopkins Press, 2005. 218 p. Cloth $40.00.

FLEISCHACKER, SAMUEL: *On Adam Smith's Wealth of Nations: A Philosophical Companion.* Princeton: University Press, 2004. xvii, 329 p. Cloth $55.00, paper $19.95.

FLIEGER, JERRY ALINE: *Is Oedipus Online? Siting Freud after Freud.* Cambridge: MIT, 2005. 332 p. Paper $19.95.

FINOCCHIARO, MAURICE A.: *Arguments about Arguments: Systematic, Critical, and Historical Essays in Logical Theory.* New York: Cambridge, 2005. ix, 467 p. Cloth $75.00, paper $31.99.

FOGELIN, ROBERT J.: *A Defense of Hume on Miracles.* Princeton: University Press, 2003. xiii, 101 p. Cloth $35.00, paper $19.95.

FOLLESDAL, DAGFINN: *Referential Opacity and Modal Logic.* New York: Routledge, 2004. xxxvi, 157 p. Cloth $65.00.

FORSTER, GREG: *John Locke's Politics of Moral Consensus.* New York: Cambridge, 2005. xi, 317 p. Cloth $75.00.

FOX, CHRIS, and SHALOM LAPPIN: *Foundations of Intensional Semantics.* Malden, MA: Blackwell, 2005. xiii, 192 p. Paper $39.95.

FRANKFURT, HARRY G.: *On Bullshit.* Princeton: University Press, 2005. 67 p. Cloth $9.95.

FRANKISH, KEITH: *Mind and Supermind.* New York: Cambridge, 2004. xiv, 255 p. Cloth $75.00.

FRANKS, PAUL W.: *All or Nothing: Systematicity, Transcendental Arguments, and Skepticism in German Idealism.* Cambridge: Harvard, 2005. viii, 440 p. Cloth $49.95.

FRASER, NANCY: *Reframing Justice.* Amsterdam: Van Gorcum, 2005. 60 p. Paper $13.74.

FREEMAN, JAMES B.: *Acceptable Premises: An Epistemic Approach to an Informal Logic Problem.* New York: Cambridge, 2005. xiv, 401 p. Cloth $85.00, paper $34.99.

FRIED, ERIC WOLF: *Inwardness and Morality.* New York: Rodopi, 2005. 155 p. Paper $40.00.

FRISCH, MATHIAS: *Inconsistency, Asymmetry, and Non-Locality: A Philosophical Investigation of Classical Electrodynamics.* New York: Oxford, 2005. viii, 212 p. Cloth $49.95.

FURTAK, RICK ANTHONY: *Wisdom in Love: Kierkegaard and the Ancient Quest for Emotional Integrity.* Notre Dame: University Press, 2005. xii, 236 p. Cloth $45.00, paper $22.00.

GABBAY, DOV M., and LARISA MAKSIMOVA: *Interpolation and Definability: Modal and Intuitionistic Logics.* New York: Oxford, 2005. xiv, 508 p. Cloth $159.50.

GALAVOTTI, MARIA CARLA: *Philosophical Introduction to Probability.* Stanford: CSLI, 2005. x, 265 p. Cloth $70.00, paper $25.00.

GALLAGHER, SHAUN: *How the Body Shapes the Mind.* New York: Oxford, 2005. 284 p. Cloth $55.00.

GALSTON, WILLIAM A.: *The Practice of Liberal Pluralism.* New York: Cambridge, 2005. x, 205 p. Cloth $65.00, paper $26.99.

GASTON, SEAN: *Derrida and Disinterest.* New York: Continuum, 2005. viii, 176 p. Cloth $105.00.

GERSHON, LLOYD P.: *Aristotle and Other Platonists.* Ithaca, NY: Cornell, 2005. ix, 335 p. Cloth $49.95.

GERT, BERNARD: *Morality: Its Nature and Justification.* New York: Oxford, 2005. xvi, 438 p. Paper $35.00.

GOLDIN, PAUL R.: *After Confucius: Studies in Early Chinese Philosophy.* Honolulu: Hawaii UP, 2005. x, 267 p. Cloth $40.00.

GROSZ, ELIZABETH: *Time Travels: Feminism, Nature, Power.* Durham, NC: Duke, 2005. viii, 257 p. Cloth $79.95, paper $22.95.

HALL, JAMES H.: *Practically Profound: Putting Philosophy to Work in Everyday Life.* Lanham, MD.: Rowman and Littlefield, 2005. xxiv, 327 p. Cloth $75.00, paper $26.95.

HAMPLE, DALE: *Arguing: Exchanging Reasons Face to Face.* Mahwah, NJ: Lawrence Erlbaum, 2005. xiii, 369 p. Cloth $99.95.

HAMPSHIRE, STUART: *Spinoza and Spinozism.* New York: Oxford, 2005. xi, 206 p. Cloth $74.00, paper $24.95.

HARMAN, GRAHAM: *Guerrilla Metaphysics: Phenomenology and the Carpentry of Things.* Chicago: Open Court, 2005. vii, 283 p. Paper $28.95.

HARMS, WILLIAMS F.: *Information and Meaning in Evolutionary Processes.* New York: Cambridge, 2004. xi, 268 p. Cloth $65.00.

HARRISON, ROBERT POQUE: *The Dominion of the Dead.* Chicago: University Press, 2003. xiii, 208 p. Paper $22.50.

HEATH, JOHN. *The Talking Greeks: Speech, Animals, and the Other in Homer, Aeschylus, and Plato.* New York: Cambridge, 2005. vii, 392 p. Cloth $90.00.

HELLER-ROAZEN, DANIEL: *Echolalias: On the Forgetting of Language.* New York: Zone, 2005. 287 p. Cloth $28.00.

HERMANN, ARNOLD: *To Think Like God: Pythagoras and Parmenides.* Las Vegas: Parmenides, 2004. xx, 374. Cloth $32.00.

HERRICK, JIM: *Humanism: An Introduction.* Amherst, NY: Prometheus, 2005. x, 105 p. Paper $20.00.

HICK, JOHN: *An Interpretation of Religion.* New Haven: Yale, 2004. xlii, 416 p. Paper $28.00.

HICKS, STEVEN R. C.: *Explaining Postmodernism.* Tempe, AZ: Scholargy, 2004. v, 230 p. Paper $18.95.

HOLDEN, THOMAS: *The Architecture of Matter.* New York: Oxford, 2004. xii, 305 p. Cloth $74.00.

HONDERICH, TED: *On Determinism and Freedom.* Edinburgh: University Press, 2005. vi, 161 p. Cloth $90.00; *On Political Means and Social Ends.* Edinburgh: University Press, 2003. vii, 182 p. Cloth $90.00; *Conservatism: Burke, Nozick, Bush, Blair?* London: Pluto, 2005. 334 p. Cloth $90.00, paper $28.95.

HÖRNQVIST, MIKAEL: *Machiavelli and Empire.* New York: Cambridge, 2004. xi, 302 p. Cloth $75.00.

HORWICH, PAUL: *From a Deflationary Point of View.* New York: Oxford, 2004. 178 p. Cloth $55.00, paper $24.95.

HUMPHREYS, PAUL: *Extending Ourselves: Computational Science, Empiricism, and Scientific Method.* New York: Oxford, 2004. x, 172 p. Cloth $47.50.

HUTTON, SARAH: *Anne Conway: A Woman Philosopher.* New York: Cambridge, 2004. viii, 271 p. Cloth $75.00.

INSOLE, CHRISTOPHER J.: *The Politics of Human Frailty: A Theological Defense of Political Liberalism.* Notre Dame: University Press, 2005. viii, 200 p. Paper $30.00.

JABLONKA, EVA, and MARION J. LAMB: *Evolution in Four Dimensions.* Cambridge: MIT, 2005. x, 462 p. Cloth $34.95.

JACOBY, RUSSELL: *Picture Imperfect: Utopian Thought for an Anti-Utopian Age.* New York: Columbia, 2005. xix, 211 p. Cloth $24.95.

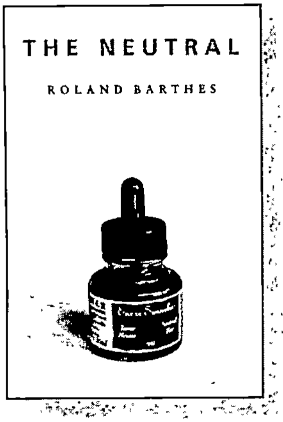
JAMES, WENDY: *The Ceremonial Animal: A New Portrait of Anthropology.* New York: Oxford, 2005. xxiii, 384 p. Paper $45.00.

JOHNSON, CURTIS N.: *Socrates and the Immoralists.* Lanham, MD: Lexington, 2005. ix, 240 p. Cloth $70.00.

KAMOLNIK, PAUL: *The Just Meritocracy.* Westport, CT: Praeger, 2005. xx, 147 p. Cloth $79.95.

KENAAN, HAGI: *The Present Personal.* New York: Columbia, 2005. xi, 199 p. Cloth $35.00.

COLUMBIA UNIVERSITY PRESS



THE NEUTRAL

ROLAND BARTHES

The Neutral
Lecture Course at the Collège de France
(1977-1978)
Roland Barthes
Translated by Rosalind Krauss and Denis Hollier

In these renowned and never-before-translated lectures, delivered in 1978 at the Collège de France, Barthes pursues a concept—the quest for the "Neutral"—that profoundly shaped his work, deepening our understanding of his intellectual itinerary and his distinctive style as a thinker and teacher.

New from Edinburgh University Press

The World, the Flesh, and the Subject
Continental Themes in Philosophy of Mind and Body
Paul Gilbert and Kathleen Lennon

This work brings contemporary continental thought into conversation with analytic philosophy on the principal topics of the philosophy of mind.

Poetics of Singularity
The Counter-Culturalist Turn in Heidegger, Derrida, Blanchot, and the later Gadamer
Timothy Clark

This book identifies and reveals the basic tenets of a forceful tradition of twentieth-century poets that differs from received ideas of deconstruction.

On Political Means and Social Ends
Ted Honderich

The moral and political arguments, judgments, and commitments of one of Britain's most outstanding radical philosophers.

On Determinism and Freedom
Ted Honderich

This volume lays out the most recalcitrant problem of philosophy—free will—and takes it beyond the standard solutions.

New from Wallflower Press

Film and Philosophy
Images of Wisdom
Adrian Page

From Plato's description of shadows on a cave wall to Heidegger's views on art, this book explores the extraordinary relationship between film and philosophy.

columbia.edu/cu/cup

CAMBRIDGE

BEST IN SCHOLARSHIP

Sextus Empiricus: Against the Logicians
Edited by Richard Bett
Cambridge Texts in the History of Philosophy
$70.00: Hardback: 0-521-82497-4: c.200pp
$25.99: Paperback: 0-521-53195-0

The Political Philosophy of Cosmopolitanism
Edited by Gillian Brock and Harry Brighouse
$70.00: Hardback: 0-521-84660-9: 272pp
$24.99: Paperback: 0-521-60909-7

The Spiritual Dimension
Religion, Philosophy and Human Value
John Cottingham
$70.00: Hardback: 0-521-84377-4: 198pp
$24.99: Paperback: 0-521-60497-4

Arguments about Arguments
Systematic, Critical, and Historical Essays in Logical Theory
Maurice A. Finocchiaro
$75.00: Hardback: 0-521-85327-3: 488pp
$31.99: Paperback: 0-521-61853-3

Paul Churchland
Edited by Brian L. Keeley
Contemporary Philosophy in Focus
$70.00: Hardback: 0-521-83011-7: 224pp
$27.99: Paperback: 0-521-53715-0

Foucault on Freedom
Johanna Oksala
Modern European Philosophy
$80.00: Hardback: 0-521-84779-6: 238pp

Aristotle's *Nicomachean Ethics*
An Introduction
Michael Pakaluk
Cambridge Introductions to Key Philosophical Texts
$70.00: Hardback: 0-521-81742-0: 358pp
$27.99: Paperback: 0-521-52068-1

Understanding Ethical Failures in Leadership
Terry L. Price
Cambridge Studies in Philosophy and Public Policy
$65.00: Hardback: 0-521-83724-3: 240pp
$24.99: Paperback: 0-521-54597-8

Punishment, Compensation, and Law
A Theory of Enforceability
Mark R. Reiff
Cambridge Studies in Philosophy and Law
$70.00: Hardback: 0-521-84669-2: 272pp

Nietzsche: *The Anti-Christ, Ecce Homo, Twilight of the Idols*
And Other Writings
Edited by Aaron Ridley and Judith Norman
Cambridge Texts in the History of Philosophy
$50.00: Hardback: 0-521-81659-9: c.320pp
$19.99: Paperback: 0-521-01688-6

Making Sense of Heritability
Neven Sesardic
Cambridge Studies in Philosophy and Biology
$80.00: Hardback: 0-521-82818-X: 281pp

Descartes Reinvented
Tom Sorell
$75.00: Hardback: 0-521-85114-9: 204pp

Mary Astell
Theorist of Freedom from Domination
Patricia Springborg
$80.00: Hardback: 0-521-84104-6: 392pp

Self to Self
Selected Essays
J. David Velleman
$70.00: Hardback: 0-521-85429-6: 400pp
$28.99: Paperback: 0-521-67024-1

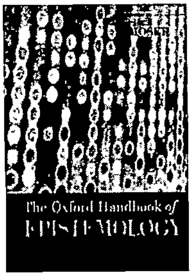
A Theory of Secession
The Case for Political Self-Determination
Christopher Heath Wellman
$70.00: Hardback: 0-521-84915-2: 218pp

Prices subject to change.

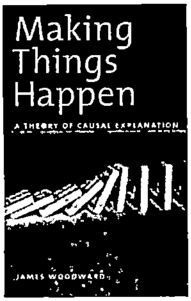
www.cambridge.org/us/philosophy/

CAMBRIDGE
UNIVERSITY PRESS

New from
Oxford



Environmental Justice
Creating Equality, Reclaiming Democracy
Kristin Shrader-Frechette



Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

New in Paperback!

The Oxford Handbook of Epistemology

Edited by PAUL K. MOSER

Written accessibly for advanced undergraduates, graduate students, and professional philosophers, the *Handbook* explains the main ideas and problems of contemporary epistemology while avoiding overly technical detail.

2002 (paper 2005) paper $35.00

New in Paperback!

Environmental Justice

Creating Equality, Reclaiming Democracy

KRISTIN SHRADER-FRECHETTE

Indigenous people, minorities, and the poor are more likely to live near toxic wastes, breathe dirty air, and work in hazardous jobs. Shrader-Frechette argues for these individuals' right to environmental justice: to equalization of the burdens of pollution, noxious development, and resource depletion.

2002 (paper 2005) paper $25.00

New in Paperback!

Making Things Happen

A Theory of Causal Explanation

JAMES WOODWARD

"*Making Things Happen* contains an elaborate presentation and defense of Woodward's manipulability theory of causation and causal explanation, a powerful alternative to extant theories in the field."—*Notre Dame Philosophical Reviews*

2003 (paper 2005) paper $24 95

Truthmakers

The Contemporary Debate

Edited by HELEN BEEBEE and JULIAN DODD

Helen Beebee and Julian Dodd have assembled a distinguished team to contribute brand new essays on the concept of truthmaking: they ask, among other things, how the truthmaker principle should be formulated, whether it is well motivated, whether it genuinely has the explanatory roles claimed for it, and whether more modest principles might serve just as well.

2005 $74.00

Thought, Reference, and Experience

Themes from the Philosophy of Gareth Evans

Edited by JOSÉ LUIS BERMÚDEZ

In this volume an outstanding international team of contributors offer illuminating perspectives on Gareth Evans's groundbreaking work, paying tribute to his achievements and leading his ideas in new directions.

October 2005 $65.00

Ramsey's Legacy

Edited by HALLVARD LILLEHAMMER and D. H. MELLOR

In these original essays, written to commemorate the centenary of Frank Ramsey's birth, a distinguished international team of contributors offer fresh perspectives on his work and show its ongoing relevance to present-day concerns.

2005 $74.00

JKC 062005

You can now subscribe to the

THE JOURNAL OF
PHILOSOPHY

using

PayPal

New Rates January 2006

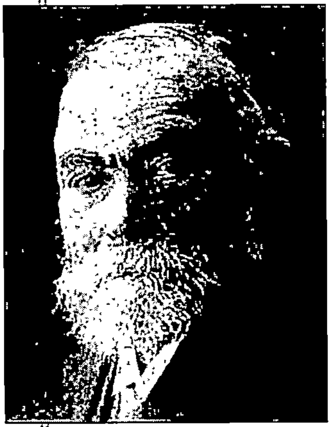
Institutional	$100
Individual	$45
Student/Retired	$20

To learn more about this easy new option, click on the PayPal icon at **www.jphil.org**

Significant

FIXING FREGE

John P. Burgess



The great logician Gottlob Frege attempted to provide a purely logical foundation for mathematics. His system collapsed when Bertrand Russell discovered a contradiction in it. Thereafter, mathematicians and logicians turned in other directions to look for a framework for modern abstract mathematics.

Fixing Frege surveys the assortment of methods put forth for fixing Frege's system, in an attempt to determine just how much of mathematics can be reconstructed in each. Here John Burgess considers every proposed fix, each with its distinctive philosophical advantages and drawbacks.

Princeton Monographs in Philosophy
Harry Frankfurt, editor
Cloth $39.95 ISBN 0-691-12231-8 Due September

New in paperback

A PHILOSOPHY OF CULTURE

The Scope of Holistic Pragmatism

Morton White

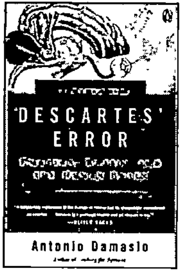
In this book, one of America's leading philosophers offers a sweeping reconsideration of the philosophy of culture in the twentieth century.

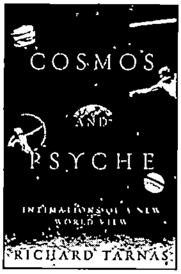
"White's picture of pragmatism is distinctive, original, and highly relevant in contemporary philosophy."
—S. Pihlström, *Transactions of the Charles S. Peirce Society*

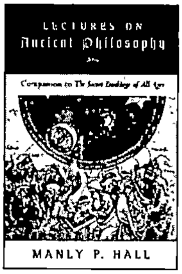
Paper $18.95 ISBN 0-691-12398-5 Due September

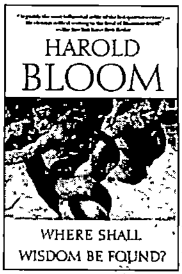
100 *Celebrating 100 Years of Excellence*
PRINCETON
University Press
800-777-4726
Read excerpts online
www.pup.princeton.edu

NEW IN PHILOSOPHY PENGUIN GROUP (USA)









DESCARTES' ERROR
Emotion, Reason, and the Human Brain
ANTONIO DAMASIO. "Crucial reading... for neuroscientists and philosophers." —*The New York Times Book Review.* "Astonishing...a passionately erudite, penetrating tour through the human mind."—*San Francisco Chronicle.*
Penguin 336 pp. 0-14-303622-X $15.00

COSMOS AND PSYCHE
Intimations of a New World View
RICHARD TARNAS. The sequel to *The Passion of the Western Mind* demonstrates the existence of a consistent correspondence between planetary movements and the unfolding drama of human history.
Viking 448 pp. 0-670-03292-1 $29.95
Available November 2005

THE RIG VEDA
Selected, Translated, and Annotated by WENDY DONIGER. Definitive translations of selections from one of the founding texts of Hindu civilization.
Penguin Classics 352 pp. 0-14-044989-2 $14.00

LECTURES ON ANCIENT PHILOSOPHY
Companion to *The Secret Teachings of All Ages*
MANLY P. HALL. Illustrated with black-and-white line art.
Tarcher 528 pp. 1-58542-432-3 $21.95

WHERE SHALL WISDOM BE FOUND?
HAROLD BLOOM. Bloom takes the reader from the Bible through the twentieth century, searching for ways literature can inform lives. "He elevates critical writing to the level of literature itself."—*The New York Times Book Review.*
Riverhead 320 pp. 1-59448-138-5 $15.00

THE ESSENTIAL TRANSCENDENTALISTS
Edited and introduced by RICHARD G. GELDARD. An anthology of core writings which provide a distinct and heretofore neglected examination of the spiritual breadth and depth of "Yankee mysticism."
Tarcher 256 pp. 1-58542-434-X $15.95

AYN RAND ANSWERS
AYN RAND. *Edited by Robert Mayhew.* A collection of Rand's Q & As from her lectures in 1957 which were intended to bring her philosophy of Objectivism to a wider audience and apply it to cultural and political issues of the time.
NAL 192 pp. 0-451-21665-2 $15.00
Available November 2005

For a FREE Philosophy brochure e-mail academic@penguin.com

www.penguin.com/academic **PENGUIN GROUP (USA)**
Academic Marketing Department 375 Hudson Street, New York, New York 10014

Balta
17.4.06

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NUMBER 10
OCTOBER 2005

page

493 *Is a Unified Description of Language-and-Thought Possible?* Joseph Almog

532 *Standing on Common Ground* Jody Azzouni and Bradley Armour-Garb

545 NEW BOOKS

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Christopher Peacocke, Philip Pettit, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.

You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2006

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.

POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2005 by the Journal of Philosophy, Inc. ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NO. 10, OCTOBER 2005

IS A UNIFIED DESCRIPTION OF LANGUAGE-AND-THOUGHT POSSIBLE?*

Frege, who started us all, was unequivocal—our account of language and thought must be unified. And since it must, Frege soon saw to it that it would. In his "Über Sinn und Bedeutung,"[1] Frege provided an elegant model of such unification. The optimism was to persist for the next eighty years. Then, around 1970, an uprising against Frege's doctrines spread like fire. The unification project ground to a halt. In what follows, I would like to (i) diagnose the reasons for the recent pessimism, and (ii) revamp the unification project.

I. THE CONTENT DILEMMA

I mentioned Frege as the originator of the unification project. In fact, Frege was the originator twice over. There was, first, his idea of unification as a regulative idea—our account of language and thought must be unified. There was, next, Frege's proposed method of unification—unification by a theory of *content*. And so, at the turn of the twentieth century, in the hands of Frege, the pursuit of the unification project took the form of what I call below the *content program*—both sentences and minds were to be assigned (thought, propositional) *contents*. The content-assignment provided in one fell swoop

---

* I owe thanks to discussions over the last fifteen years with Tony Anderson, Keith Donnellan, Kit Fine, Stavroula Glezakos, Paolo Leonardi, Sten Lindstrom, and more recently, Roberta Ballarin, John Carriero, Bob Stalnaker, and Ori Simchen. Special thanks are due to Howie Wettstein and Erin Eaker whose ideas have shaped mine (see references below). My heaviest debt is to David Kaplan who served both as chief target (see below, the "disunity thesis") and chief helper. A footnote tracing the long history of the paper is attached at its very end.

[1] In what follows I refer to "Über Sinn und Bedeutung," as *SB*. I rely on my preferred translation, "On Sense and Nominatum," in Herbert Feigl and Wilfrid Sellars, eds., *Readings in Philosophical Analysis* (New York: Appleton-Century-Crofts, 1949), pp. 85–102.

0022-362X/05/0210/493–531

(i) *semantical* descriptions of public language sentences like "London is pretty"—what the sentence means, and (ii) *psychological* descriptions of thinkers—what a certain man, Pierre, thinks, when he thinks that-London-is-pretty.

The unification project reigned until the emergence, around 1970 (a fateful date to which we shall return), of those I will refer to as the "quartet," namely, Keith Donnellan, Saul Kripke, Hilary Putnam, and David Kaplan. Whereas the founding fathers, Frege and Russell, were attentive to (the hyphened) Language-and-Thought, the quartet is often presented as focused *exclusively* on language. The quartet engendered a set of ideas, importantly not a *theory*, that is often called "Millian" or "direct reference" semantics. I shall rather call the semantics, following Donnellan, a *purely referential semantics* (reasons for insisting on this new nomenclature—as opposed to the common "direct reference" or "Millianism"—will emerge in due course).

With both the original content program and the latter-day purely referential semantics on board, we have the incompatible ingredients leading to the undoing of the unification project. The incompatibility is encapsulated by the *content dilemma.*

One horn of the dilemma is engendered by the Frege (and in his wake, Russell) methodology. Driven by *psychological* data we are about to review, Frege-Russell posit rich contents to describe both (i) our thoughts about objects, and (ii) the semantics of expressions we use to refer to those objects. But, alas, the 1970 data of the quartet seems to show that the psychologically driven contents mis-describe the semantics of our words, for example, proper and common nouns like 'Cicero' and 'Elm(s)' (in the case of this last both in its singular and bare-plural deployments). The Frege-Russell contents give a false account of (i) the *mechanism* by which the nouns reach their referents, (ii) what we *say* when we use the nouns, and (iii) what we *know* (have "in the head") when we use the nouns. This suggests we should rid ourselves of the Frege-Russell contents.

Can we? Shedding the contents sends us from the frying pan into the fire. We need only recall the original psychological data that got Frege-Russell going. Our intuitive descriptions of the psychology of ordinary thinkers entrap the purely referential and content-free new semantical account with puzzles. For our intuitive descriptions require us to distinguish in a thinker like Pierre, for example, the thought that Cicero was an orator (respectively: that London is pretty; that elms are green) from the thought that Tully was an orator (respectively, que Londres est jolie; que les ormes sont vert). And putting aside thoughts, which after all are rather abstract, there is surely direct psychological reality to Frege's more tangible notions of what is "cognitively significant" and "informative." Thus the mathematician Lagrange (and a bit after him, Gauss) found it startlingly informative to learn that Ms. Sophie Germain (a woman Lagrange met at the grocer shop in the Paris 14th district) is one and the same as the *soi-disant* Monsieur Leblanc (a clever student of prime numbers whose correspondence he was studying with great interest). "Sophie Germain=Monsieur Leblanc" was surely as cognitively significant and informative for these distinguished agents as things ever get.

Yet these natural observations about cognitive significance, informativeness, and thought identity seem not available to the content-impoverished purely referential semantics. And so, the pressure to choose between the purely referential and the content-theoretical semantics traps us between hammer and anvil—too little content for psychology, too much content for semantics. Thus, the *content dilemma.* As if the aforementioned is not tension enough between post-Fregean data and Frege's original unified theory, I should like it noted that Frege himself "smelled" the tension I am highlighting. Having just argued for denotation-transcending contents (his *senses*), Frege footnotes his discussion with a new ticklish piece of data. In the second footnote of *SB*, Frege says:

> In the case of an actual proper name such as 'Aristotle' opinions as to the sense may differ. It might for instance be taken to be the following: "the pupil of Plato and teacher of Alexander the Great." Anybody who does this will attach another sense to the sentence "Aristotle was born in Stagira" than will a man who takes as sense of the name: "the teacher of Alexander the Great who was born in Stagira." So long as the denotation remains the same, such variations of sense may be tolerated... (*op. cit.*, p. 86).[2]

We all read Frege's footnote, comment it in passing, and go on to view it as not impinging on the prospects of his unification project; at least I did for many years. But it is anything but tangential. For in the footnote, Frege brings out two *facts* of great significance for both his own theory, as for any other theory: (i) the *variability of speaker's associations,* and (ii) the seeming inescapability of indirect discourse *reporting by sheer-denotation.*

Frege himself, having noticed the two facts, tells us, rather nonchalantly, that "so long as the denotation remains the same, such variations of sense may be tolerated...." Tolerated? How could he use of all adjectives this one? First, given how our cognitive lives work—

---

[2] I have made several alterations in the translation and used (Church's) "denotation" instead of the translators' "nominatum."

how language is propagated and how scattered in space and time thinkers are—both facts, the variability-of-association and reporting-by-sheer-denotation, are anything but accidental. How exactly does Frege plan to suppress the two phenomena? And if they are irrepressible, what does it mean to tolerate such structural facts about our cognitive lives? We may as well say that we are going to tolerate the fact that information travels no faster than light or that our brains are made of physical particles. We do not "tolerate" such facts; we live by them.

Second, tolerate these facts as we may (have to), I cannot see how Frege can even fantasize about a *unified* theory of language and thought as long as he himself emphasizes to us their drifting apart. For on the one hand, as his footnote just recorded, (i) the *linguistic* phenomena—uses of proper names to refer, communicate, and report attitudes—may well be sheer-denotation bound. On the other hand, when it comes to (ii) *cognition* proper, to the *thoughts* you and I have, denotation-transcending contents still dominate. Each of our (respective) thoughts is—and essentially so—content-made after all. Was then Frege, in spite of his unification ideal, an early discoverer of a *dis*-unity thesis—language is denotation-bound, thought content-bound?[3]

---

[3] In oral presentations, some (though not all) connoisseurs of Frege's footnote and his official theory of indirect discourse made to me at this point the following (partial) objection. They pointed out that (i) Frege completes the footnote with the remark that the aforementioned *facts* afflict ordinary ("actual") language but no such situation is to be allowed in a "language of demonstrative science," (ii) Frege had an official shifty theory of "intensional" indirect discourse (to recall—the words in the complement clause do not denote anymore the ordinary denotation (the man, Aristotle) but *the* sense of the name (one feels like asking Frege and his defenders: *The* sense? *Which* one?)). His official theory would suggest that he would "tolerate" the operation by denotation-only solely in extensional contexts.

My quick response is focused at this point not on defending any theoretical points (this is left for developments below) but mainly on emphasizing what indomitable *brute fact* had been unearthed here by Frege.

First, since our interest here is in the "actual" ordinary English in which I report Pierre, let us leave aside Frege's stipulation about a "language of demonstrative science." (I will say telegraphically this much about the latter, however. Let Pierre and myself be demonstrative scientists. Let Frege's language of demonstrative science include locutions like "Pierre believes that...." Pierre has been thinking ("scientifically") about Aristotle and I want to report his thought to a scientific congress. Does Frege have some way to get inside Pierre's and my mind and force us (over time!) to associate the same sense with a given formal Aristotle-denoting individual constant '*A*'? (i) I do not think Frege could pull this off simply because I do not think even God can go inside our minds and so force us (as long we deem impossible (I do) our all thinking of Aristotle by means of a fixed "Aristotle-chip," in the manner of the Star Trek character Data); (ii) even in the best case scenario—for example, thought about numbers, using standard designators like '144169'—I do not see how I could be forced to think of the number as '1 times 100,000 + 4 times 10,000 + 4 times 1000 + 1 time 100 + 6 times 10 + 9'. Will my report to the number theory congress that "Pierre believed for

II. BASIC METHODOLOGY: REFERENCE, CONTENT, INFORMATIVENESS—FACT VERSUS THEORY

Before we move on to recent responses to the content dilemma, I would like to sensitize our discussion of key notions like "(direct) reference," "content," and "informativeness" to a distinction I keep emphasizing, between pre-theoretic *facts* ("data") and *theoretical* conjectures.

In both our understanding of Frege and of the quartet, it seems that we often read into the sheer data ("facts") spotted by these authors, some of the authors' conjectures and underlying theories. I give examples relating to both Frege's work and later work by the quartet. The examples bring out the data/theory distinction in connection to both semantic and psychological questions.

II.1. Frege's Puzzle of Simple Cognitions: Fact versus Theory. I will rely below on a distinction between questions about *simple cognitions* and *reporting cognitions.* The former cluster of issues concerns our analysis—so to speak "from the outside" or "in the metalanguage"—of what it takes for an agent, say the man Pierre, to stand in a cognitive relation (think of, believe about, and so on) to a certain item, for example, the city of London. By the latter set of issues I target our ways of

---

a long time and then proved that the square root of 144169 had some special features" not be accepted as true because my sense for '144169' is not Pierre's?)

Now demonstrative science aside, in the case of ordinary language and actual names like 'Aristotle', Frege cannot simply *stipulate* that I use 'Aristotle' with the sense Pierre had in mind (when he said "Aristotle was clever"). Notice that I am not here questioning Frege about the *cognitive facts* (Do we—Pierre and myself—have each a uniquely identifying sense for Aristotle?). At this point, I am stipulating that Frege has the cognitive facts right and I am asking a question about the *language of reporting.* How can I report what is in Pierre's head?

I could of course use an existential quantifier over senses (though notice we do not ever do so in ordinary English) and say: "there exists a certain sense *x* in Pierre's head, heaven knows which one it is (I do not) and there is a sense *y* in my head and they denote one and the same individual, namely, Aristotle, and...." Given our stipulations of the cognitive facts, this is no doubt *true of* Pierre but it is no reporting of his *specific*-Aristotle thought. What else could I do?

My (and I submit Frege's) main point in the text antecedent to this note (and further developed below) is that independently of what I (reporters) *could* do, what we all *actually* do is use indirect discourse with an actual proper name, 'Aristotle' (and assuming again that for me as for Pierre the word connotes a distinct sense). We use this actual name—with varying senses—whether the "context" is the extensional "it is not the case that..." or the intensional "Pierre believes that...." And we just go on to *tolerate* the fact that I will not get at the sense Pierre thinks-with; I will only get at the shared denotation. This much is a *brute* fact about what we actually do in reporting other people's beliefs.

Quite apart of his *theories* and *ideals,* Frege was first and foremost a remarkable philosopher *by ear.* He knew that this is a brute fact about our use of actual language and he smelled that it is serious trouble for his account. I would like to see where this fact takes us.

reporting-in-language such cognitions. So here we are concerned with the semantics of object-language sentences, for example, "Pierre believes that London is pretty" or "Pierre believes London to be pretty," reporting about Pierre's cognitive life.

What we call "Frege's puzzle" concerns a set of issues about simple cognitions. In stating the puzzle, we often run together (i) Frege's sheer *datum*, for example, the sentence "Cicero=Tully" is informative (cognitively significant) for a given agent, for example, Pierre, with (ii) Frege's conjectured theoretical *explanation* of cognitive processing. There is first Frege's explanation of (a) *the cognitive bond*—what makes Pierre's thought about this ancient Roman man, using either 'Cicero' or 'Tully', possible. Frege claims that Pierre associates with each of the names a representational content (Frege: "sense") that uniquely identifies in the world that man. Next, (b) Frege provides an explanation of *the cognitive significance* of "*a*=*b*" identifications, for example, the identification "Cicero=Tully" is *informative* (cognitively significant). To explain this last, Frege asserts that not only does Pierre have in the head *some* content when he thinks with 'Cicero' and *some* content when he thinks with 'Tully'; he must have *distinct* such contents. It is the distinctness of the contents that generates the informativeness (cognitive significance) of the identity "Cicero=Tully."

I propose a more prudent methodology that disengages fact and theoretical conjecture. The *datum* proper is (i)—the identity "Cicero=Tully" is informative (cognitively significant). Let us dub this *Frege's informativeness fact*. Frege's two-pronged explanation of (a) the cognitive bond—thought of Cicero (Tully) requires the *existence* of a mediating content and his explanation of (b) the informativeness of the identity—it consists in (or: arises from) two *distinct* information pieces associated with 'Cicero' and 'Tully', asserts theoretical *conjectures* about how the cognition of the object (and the cognitive significance of the identity) come about. I will call the first conjecture Frege's *existence (of content) conjecture*:

> (Existence) No cognition without representation (content).

Frege's second conjecture deserves the title of Frege's *distinctness (of content) conjecture*:

> (Distinctness) "*a*=*b*" is informative iff distinct contents are associated with '*a*' and '*b*'.[4]

---

[4] I emphasize here the general *logical structure* of Frege's conjectures, rather than the specific kind of contents or senses (are they descriptive? Purely qualitative? Expressible by a definite description? And so on) the historical Frege plugged into the conjectures. Neo-Fregeans may (did) question Frege's specific kinds of senses (con-

II.2. Frege's Footnote: Fact versus Theory. Many discussions of Frege's footnote run together facts recorded in the footnote with (Frege's own) favorite theoretical explanations of the facts. Thus it is often said that we see directly in Frege's note that our use of the public language (to refer, to report attitudes) is sheer denotation-bound but individual-cognition (thought) is denotation-transcendent and essentially content-bound. It is also often suggested that my indirect discourse report "You believe that Aristotle was clever" is not strictly speaking true because (i) you and I have different thought contents, and (ii) I have not reproduced yours.

I urge here a sharp separation of fact and theory. It seems to me that in the footnote, Frege makes us aware of certain *observational* facts (whether he himself distills those facts and separates them from his own theoretical agenda). The first such fact is the *variability of associated-information.* Different people, for example, you and me, or the same person, myself, at different times, may respond with different (definite) descriptions to the question "So, who was Aristotle?" This *is* a fact. Whether the description associated articulates (i) the *meaning* of the term used and (ii) the *thought content* the agent is thinking are theoretical questions. It may well turn out that this information is merely a *communication helper* offered to the audience so that it could *identify* who Aristotle is (was), but with both my linguistic reference to and thought of that ancient Greek fixed antecedently by quite a different mechanism (rather in the way that my pointing finger, when I use the word 'this', may well help you identify the demonstratum; nonetheless, it is a different and prior mechanism that fixes what I refer to and who it is I am thinking of). To reiterate, that we respond with different information pieces is a fact; what this information does (where it fits in the theory) is a much more complex theoretical issue that requires some analysis.

In the same vein, I take it to be a pre-theoretic *fact* that we all use indirect discourse to report truly the likes of "John believes that Aristotle was clever and so does George," "I used to believe Aristotle was clever and I still believe he was." Quite a different level of discussion is involved when we wonder: What does this reporting by sheer denotation show about the *metaphysics* of our cognitive states proper, about what it is for you (me- through- time) to think of Aristotle (for example, that he was clever)?

---

tents) but preserved his general conjectures. I am here focused on discussing the conjectures at this level of generality. My formulation of the existence conjecture is a variant of one due to Kaplan ("no mentation without representation").

In sum, I would like to segregate facts about truth values of indirect discourse sentences from conjectures about the role of the information we provide when we ask “So, who is Aristotle?” or “So, what is London?” The contemporary received view, often stated without argument, that these associated descriptions are our very thought contents (washed away by our vernacular reporting) is a theoretical claim, not a fact. We shall have to consider below whether these descriptions are indeed constitutive to the very identity of the cognitive states.

II.3. Referring: Fact versus Theory. A basic tenet of the 1970s work on the semantics of referential devices (by Donnelllan, Kripke, Kaplan, and Putnam) is that (proper, common) names and indexical pronouns are purely referential devices. Kripke has argued that the name ‘Nixon’ differs from descriptions like “the US president in 1970” vis-à-vis modal evaluations. Donnellan has pointed out that for any such definite description ‘the *F*’, indeed even one that is “modally rigid” by Kripke’s lights (“the offspring of such and such gametes”), it *could turn out* that Nixon is not the *F* but of course it could not turn out that Nixon is not Nixon. Third, Putnam observed that expressions like ‘Elm’ and ‘Beech’ behave very much like proper names and that descriptive definitions ‘the *F*’ of the purported tree species are (i) not available to the lay user, and (ii) any such description ‘the *F*’, phenomenal or scientific, could turn out wrong of elms and yet they would still be elms. Fourth and final, Kaplan pointed out that indexicals and demonstratives expressions, like ‘I’ and ‘he’, are used differently than any descriptive phrase, including phrases like “the agent of the context” or “the object demonstrated in the context.” No such descriptive phrase is a synonym of the indexical and the demonstrative.

The foregoing are pre-theoretic facts—arguments “by ear”—about our use of (proper, common) names and indexical pronouns.

It has become commonplace to state these direct observational facts in terms of certain familiar theoretical frameworks of (possible worlds, propositional) semantics. Let me focus in this respect on Kaplan’s idea of direct reference and leave the notion of modal rigidity aside.[5] A standard view in the contemporary literature is Kaplan’s popular *hypothesis* that sentences express propositions, be they singular or general, and that names (indexicals) are distinctive in having their referent determine their propositional content (rather than the other way round). About this hypothesis, I suggest the following methodology. It may be that such a *propositional*

---

[5] Whatever I have to say about the noninnocent assumptions behind the theoretical articulation of “direct reference” applies with bells on to “modal rigidity.”

(*content*) semantics is the most natural theoretical way of explaining the observational "by ear" arguments. But again we need to proceed here with caution, for (i) whereas Kaplan articulated the intuitive idea of pure reference by means of singular propositional contents, Kripke and Donnellan, surely as tenacious advocates of pure reference, have refrained from calling upon any propositional semantics to explain that 'Nixon' (and 'Santa Claus') are purely referential, and (ii) it may well turn out that there are cases where the propositional semantics apparatus breaks down but the names are as purely referential as ever. Donnellan was the first to point out one such example, empty names. And it may well turn out that there are other such examples. For example, in summing up his "A Puzzle about Belief" paper,[6] Kripke speaks of Pierre-like indirect discourse reports as witnessing the "breakdown" of the notions of *content* and *proposition.* Suppose this were the case—such reports display the inapplicability of the apparatus of propositional contents. Would this show that 'London' is not purely referential?[7]

In speaking of reference, we need to separate fact and theoretical conjecture. I will take it below that the quartet's work has pointed out the pure referentiality of names and indexicals. Only one of the four, Kaplan, has gone on to offer a content-theoretical semantics, indeed a two-tiered amplification, with (i) character ("meaning") and (ii) propositional content, offering revisions and generalizations of the Frege (and Russell) ideas of (i) linguistic meaning and (ii) thought content. This generalized content semantics is not part of what I take to be the sheer (linguistic, semantic) data pointed by the quartet. It is a conjecture open to refutation.

---

[6] In Avishai Margalit, ed., *Meaning and Use* (Boston: Reidel, 1979), pp. 239–83.

[7] Donnellan writes about empty names in "Speaking of Nothing," *Philosophical Review*, LXXXIII (1974): 3–32. Kripke himself points out that no propositional contents are generated by empty names in his unpublished John Locke lectures (Oxford 1973), lecture VI. Donnellan commented earlier on the breakdown of propositional semantics in discussing "referential" descriptions and what "statement" such uses produce. I have discussed Donnellan's and other cases of insufficiency of propositional semantics, first in "The Subject-predicate Class I," *Nous*, XXV (1991): 591–619, and later on in "The Proper Form of Semantics," in Marga Reimer and Anne Bezuidenhout, eds., *Descriptions and Beyond* (New York: Oxford, 2004), pp. 390–419. It should also be clear by now why in describing the theory-free observations of the quartet, I prefer not to use the popular titles of "Millian" and "directly referential" semantics. Both of these titles are infused with theory. We have just seen that "directly referential," as introduced by Kaplan, is stated in terms of contributions to "content"; what is more, the Kaplan paradigm of a directly referential term, indexicals, are further associated with a rich character-meaning, a second level of representational content. This has led Kaplan to look for Neo-Fregean solutions, using characters, to informative identities of the form "I am he" (on which more below). The notion of a Millian expression is slightly less theoretical and has been introduced by Kripke for an expression whose "sole linguistic function is exhausted by its standing for its bearer." But even this involves theoretical conjectures (are empty, bearer-less, names disfunctional?); and even for functional names, there is in "*linguistic* function," a conjecture about there not being any linguistic contribution of the word proper beyond its *semantic* value. What is more, the title "Millian" is somewhat a misnomer for my purposes. Mill himself thought that (i) other single word nouns, for example, common nouns like 'cow', are connotative. This would make the Kripke-Putnam name like treatment of such kind-naming nouns...non-Millian, a confusing nomenclature. (ii) It seems that Mill himself thought that in identities like "Cicero=Tully" we speak, not of this one orator, but of his two names. In contrast, Donnellan's term "purely referential" involves (i) no theory behind it, and (ii) means just what the two English words it uses mean in the vernacular.

III. REACTIONS TO THE 1970 TURN—MORE CONTENT VERSUS NO CONTENT

I would like to compare the reaction of two general methodologies to the quartet's 1970 data—(i) the *semantic reformation* and (ii) the *cognitive reformation*.

The semantic reformation methodology, working from *inside* the content program, reads both the quartet's data and the cognitive puzzles as (i) confirming the correctness of the Frege-Russell *cognitive* account—the analysis of how we *think*, and (ii) calling for a revision of our *semantic* theory—we need to amend the Frege-Russell account of how we *refer* (in language). The revision calls for the generation of more *kinds* of thought contents, more world-involving ("singular," "existence dependent") contents than those admitted by the pure propositions (*gedanken*) of Frege.

The cognitive reformation methodology offers an orthogonal diagnosis: what we need is not a richer dose of the troublemaker notion of *content* (proposition) but exactly the opposite—an account that is through and through *content-free*. What is more, the cognitive puzzles are not a confirmation of the Frege-Russell approach to cognition but, again, quite the opposite: the primal error of Frege and Russell is in the very realm most of us take them to have their strongest card—the analysis of *how we think*. The quartet's 1970 data demands that we rethink the very role of contents in describing our cognitive lives.

III.1. Reformation Methodology I: More Content. Many responses of the "more content" variety have been given since 1970. Some are minor revisions of the classical Frege-Russell framework, some more general overhauls. To give us a taste of such reactions, I review central Neo-Fregean and Neo-Russellian reactions. I am not here bent on evaluating the responses so much as bringing out, as it were "from the inside," the motivating idea behind each response.

A. Grade I Revision: A Mixed Mill-Frege Account. Already in his 1978 "A Puzzle" piece, Kripke reviews a response-reformation of Frege, in light of the 1970 troublesome data. He calls it the "mixed view," for indeed it mixes some aspects of Frege, with some ideas of Mill (and

Kripke). The mixed view introduces amendments regarding the behavior of names in modal contexts but keeps the Frege-Russell cognitive idea of a full fledged identificatory (descriptive) sense.[8] Consider the (lore's) Frege-Russell descriptions for 'Cicero' and 'Tully', for example, "the denouncer of Catiline" and "the author of De Fato." On the mixed view, we rigidify them by means of the adjective 'actual'. We now have on hand "the actual denouncer of Catiline" and "the actual author of De Fato." A variant solution assigns the distinct (rigidified) "metalinguistic" descriptions "the actual referent of 'Cicero'" and "the actual referent of 'Tully'." Either way, the contents thus introduced by 'Cicero' and 'Tully' into thoughts are *altered* contents, each a modally rigid *sinn*.[9]

However when it comes to the cognitive puzzles, we reapply the original Frege-Russell methodology. "Cicero=Tully" is said to be informative because it expresses a thought whose content is informative, the thought that *the actual denouncer of Catiline is the actual author of De Fato* (or possibly: *the actual referent of 'Cicero' is the actual referent of 'Tully'*). Similarly, the troubling indirect discourse report "Pierre believes that Cicero was an orator but that Tully was not" is vindicated as true and of a perfectly rational Pierre by reading it as ascribing a noncontradictory content to Pierre: that the actual denouncer of Catiline was an orator but that the actual author of De Fato was not (or: that the actual referent of 'Cicero' was an orator but the actual referent of 'Tully' was not).

We encounter in this early "more content" response the pattern that will recur below in later sophisticated variants. The "more content" methodology is to secure by means of a (revised) content assignment what I will call the *twofold adequacy condition*:

(i) The *psychological adequacy condition*—reflect the psychological data provided by Frege's and indirect discourse puzzles.
(ii) The *semantic adequacy condition*—reflect the correct truth conditions of (directly referential) singular terms as displayed by their behavior in modal and temporal embeddings.

B. Grade II Revision: Two Levels (Dimensions) of Content. The aforementioned "mixed view" indulges in a brute force retrofitting of the Frege-Russell semantics. A *single* level of content, for example, of a proper name like 'Cicero'—is made to satisfy both adequacy conditions. More far-reaching reformations of the Frege-Russell structure have been proposed in the last three decades by various frameworks. The leading idea has been to split the notion of content in two levels (dimensions, and so on)—to each adequacy condition, its own kind of content. The variations on the two-dimensional content theme are many. I will discuss here in some detail the subsequent suggestions of Kaplan in the last thirty years.[10]

---

[8] See "A Puzzle about Belief." I refer to Kripke's discussion in section I of what is called the "mixed view."

[9] Analog *temporal* rigidification may be secured by means of the adjective 'present' or the modifier 'at *t*', so that the purported *sinn* picks Cicero (=Tully) at all times.

C. Cognitive Description by Character. Consider my true assertion of "I am he" (with me realizing, as I look in the mirror, the awful looking face is mine). The semantic content is the identity singular proposition that Joseph Almog is Joseph Almog. But the *sentence* is informative. To explain the informativeness, Kaplan calls upon (i) the *conventional* meanings of the two personal pronouns, "the agent of the context" and "the male demonstrated in the context," (ii) two meanings that are both said to be *cognitively* available, "in the head" of the competent user of such pronouns. Kaplan calls such meanings *characters*. Kaplan now applies the above-mentioned Fregean *distinctness of content* conjecture: it is two distinct information pieces that are in the head of the user that induce the informativeness of the whole sentence.

The proposal rests its case on explaining informativeness by conventional linguistic meanings—Kaplan's "characters." This leaves Kaplan with three major problems.

Proper Names. Proper names do not seem to have such rich descriptive meanings. The sole linguistic function of 'Cicero' is exhausted by its standing for its bearer and the very same is true of

---

[10] The idea to use two dimensions of content, in one sense or another, recurs in many post-1970 responses. I have elected to discuss Kaplan's work for three related reasons. (In extended oral presentations of the present material, I have analyzed in this vein, Robert Stalnaker's two-dimensional modal semantics, Bas van Fraassen's early and cunning but little known application of it to belief reports, Gareth Evans's theory of type-token modes of presentations, and John Perry's account of the semantic and cognitive values of indexicals. References to the aforementioned frameworks are added below when pertinent.) The first reason to focus on Kaplan's framework is that of all the Frege-Russell revisions I am familiar with, Kaplan's is the most radical on both the semantic and cognitive fronts. If the diagnoses below of the limits of Kaplan's framework are valid, they apply *inter alia* to more conservative revisions. A second related reason is that Kaplan takes himself to revise *Russell's* propositional semantics. Kaplan thus allows, right from the start, more types of contents ("singular propositions") to be assigned by the simple sentence semantics. Again, if even this kind of generalization fails, this leads us to suspect something in the very idea of a content semantics is problematic. Third and still interrelated, Kaplan's work has gone in the last thirty years through subsequent *revisions* and developments reflecting just the kind of worries about the limits of content semantics I have in mind. In examining the pressure for the revisions, we can bring out the inherent limits of the "more content" methodology.

'Tully'. If characters were articulated for such names, the two names would have the same character.[11]

Are Characters in the Head? Indexicals (demonstratives) are associated, by convention, with lexical meanings. But it is far from clear that these rule-like "meanings" are cognitively available, are in the head of the lay user. The use of the expression may well be describable by such a rule, discoverable by a theoretician surveying our tribal practice "from above." But this is far from ordinary practitioners "having," to use Russell's pregnant phrase, the character. The articulation of characters seems rather like that of historical chains of transmission, discovered by Donnellan and Kripke—the mechanism may well explain the functioning of a proper name across history. But is it part of the notions (concepts, ideas, information) carried by an ordinary user of the name 'Cicero'? Donnellan submits that the articulation of such a chain-mechanism belongs not with the mundane users of names but is the work of an "omniscient observer of history," overseeing our practices "from above." In recent work, Kaplan seems to consign the articulation of characters ("rules of use") to this level. Attention to the complexity involved in articulating the rules for demonstratives brings out just how controversial (and theoretical) is the matter of describing the laws ("rules") governing our practice of the word 'this'. And even the indexical 'I' is not easy to describe. We may know each, in a special way, the truths "I exist," "I think," and other such expressions of self-knowledge. But this is not to say ordinary users would come up with Kaplan's rule, namely, "the agent of the context," rather than, for example, "the speaker (utterer)."[12]

Indirect Discourse Embeddings. The previous problem arose from the cognitive unavailability of the conventional meanings. A starker form of cognitive unavailability, known to Kaplan all the way back in

---

[11] In the formal set up of Kaplan's model theory of his "Logic of Demonstratives," one could, in the manner of Stalnaker's work, assign (modulo set theoretic problems of domain "size") the two expressions, functions from model structures to characters (linguistic meanings). "Cicero=Tully" is thus made "logically contingent," true in some (our "intended") model structure, false in others. But, as Kaplan correctly insists, *in the English language as we use it,* if the notions of (i) meaning and (ii) character apply to proper names at all, the co-referring names get the same character-meaning and "Cicero=Tully" is true in virtue of character alone.

[12] Kaplan is focused on evaluations of sentences with 'I' in which no actual physical use (speech) of the pronoun has to occur (for example, I am asleep). The pushing up of the characters from the head of the thinker(user) to the omniscient observer level is broached in Kaplan's as yet unpublished "What Is Meaning?" That this is the right level at which to articulate characters was suggested some time back in my own "Semantical Anthropology," *Midwest Studies in Philosophy,* x (1984): 479–90. A similar "meta-semantical" interpretation of the two-dimensional formal semantics has been embraced lately by Stalnaker in his "Assertion Revisited," *Philosophical Studies,* CXVIII (2004): 299–322.

Demonstratives,[13] shows up when we consider indirect discourse reports, with indexicals (demonstratives) in the embedded clause.

Suppose I say to David Kaplan in his L.A. office "Pierre believes that you are here." Surely this is true not because Pierre said to himself ("silently") back in Paris "you are here." Pierre may have said back in Paris to his friends, "Le professeur qui nous a parlait la semaine dernière avec ce costume élégant de Saville Row est de retour dans la cité des anges." The verdict of the omniscient observer of history, checking "from above" on (i) Pierre's Parisian mental (verbal) episode, and (ii) my Los Angeles verbal report, will be that I got Pierre right, even if the two of us, reporter and reported, are quite unaware of one another. What then explains the truth of my report "Pierre believes that you are here"? My embedded clause matches neither in character nor even in content (singular proposition) what Pierre says back in Paris. Of course, Frege could whisper for us from his *SB* footnote that "I predicted all this. For as long as the denotation is the same...." But the concession, by Frege as by Kaplan, only raises more questions: What kind of content-theoretical foundations of semantics and cognition are we provided if what saves the day and grounds the report is the content-free sameness of denotation?

D. Grade III Revision (1977): Indirect Discourse Reports «from Above». In 1977, Kaplan noticed that such reports with embedded indexicals, "Pierre said (believed) that you are here," signify the breakdown of the classical Fregean strategy of taking the semantical unity—the meaning—generated of the embedded clause, for example, "you are here," and making the indirect discourse verb operate on it. Indirect discourse is not direct discourse—the reporter is not quoting the thinker; the reporter is using his own words in the embedded clause.

In *Demonstratives*, Kaplan proposed that characters are *mentioned* in the *meta*-language, in stating "from the outside" the truth conditions of my object language indirect discourse report. Thus:

> (KAP) "Pierre said that I was you" is true in context *c* iff there is a context *c** with Pierre as the agent and there exist words *A* and *B* with characters ch-1 and ch-2 such that when ch-1 and ch-2 are set in *c** they pick the same objects-contents as the characters of 'I' and 'you' in *c*, and at *c**, Pierre assented to "*A* is *B*."

In (KAP), characters are mentioned as part of the background facts that make my indirect discourse true. They are not part of what I report Pierre as believing, the alleged content of his belief. And so, we have drifted away from the content program: the embedded clause—if, an important "if," it has indexicals in it—is not the reproduction, by a passive reporter, of the state the thinker's mind is in. Indirect discourse reporting is rather the work of an active reporter, getting at the very items thought of by the thinker.[14] With hindsight, I would say that Kaplan neo-Fregean theory of 1977 aimed to preserve as much as possible of the content program. This was patently not possible when the reporter used indexicals. Barring the use of direct discourse quotes, the reporter cannot reproduce (verbatim) the manner in which the thinker thinks of the referent. But as long as the term used by the reporter was context-free, we may insist that the function of the report was the passive reproduction—in the embedded clause—of how the thinker was thinking of the target object. In all, in the 1977 theory, the function of the report was the reproduction of the thinker's mind; the case of indexicals was viewed an aberration that subverts this function.[15]

---

[13] Published in Almog, Perry, and Wettstein, eds., *Themes from Kaplan* (New York: Oxford, 1989), pp. 481–564.

E. Grade IV Revision (2001): The Disunity-thesis: De Re Reports, De Dicto Beliefs.[16] The term-sensitivity of the very function of the report should strike us as somewhat capricious. And indeed, two decades later,

---

[14] Another writer who realized at the time (1977) the need to read the indirect discourse semantics "from above" was van Fraassen, in his little known but beautifully laid out piece, "Propositional Attitudes in Weak Pragmatics," *Studia Logica*, XXXVIII (1979): 365–74. van Fraassen goes on to anticipate there Kaplan's eventual separation (see below) of (i) the semantics of reports of belief, and (ii) the metaphysical analysis of the structure of the beliefs proper.

[15] As mentioned, this is very much Frege's own method both in the "Aristotle"-footnote of *SB* and in his later article "The Thought." Indexicals and what Frege calls "actual" proper names are identified as (i) hard cases that (ii) make a bad law, for the otherwise content-bound indirect discourse. Hard cases indeed: if Gustav Lauben says "I am wounded," how can I report him except by "Gustav said that he (Gustav) is wounded"? For Frege, this does not, stronger yet *cannot*, preserve the original content-sinn; if Gustav says last Monday, "Today is my birthday," how can I report him the next Monday except by the likes of "Gustav said that last Monday was his birthday"? If he said on that Monday a week ago, "Next Monday is my birthday" would I not report him right by saying a week later "Gustav said that today is (would be) his birthday"?)

More generally, in weighing the hard cases versus the desired laws, I feel that both Frege and Kaplan are ambivalent between (i) what they *hear* by listening to the use of our language and (ii) their *theoretical* inclinations. Both Frege and Kaplan noticed (and Kaplan provided in the above (KAP) a *mechanism* that allowed) the possibility of treating in content-free ways indirect discourse reports. Indeed, I feel both inspire us to contemplate the stronger result that indirect discourse *must* be so treated (other such acute *examples*, see below, have been provided by Quine, Donnellan, and Russell's yacht case). In the end, Frege and Kaplan follow their content-imbued theoretical predispositions. But, in my view, it is their good ear that inspires us to question the content program.

[16] In this section, I owe thanks to discussions with Eaker.

Kaplan revisits the question in his recent "De Re Belief."[17] Under the impulse of his student Erin Eaker and his own arguments by "ear," Kaplan has come to recognize the pervasiveness of what was originally just a little trickle ("indexicals") in the Fregean dam. The critical point to recognize is not only that the content-program context-free "safe terms" (proper names, definite descriptions) are more sheer-denotation bound than the program allowed. The key observation is rather that the whole issue does not turn on which terms are used but rather on who uses them. Kaplan's new focus is the unstoppable "auto-pilot" dynamics of reporting, whatever singular terms are used.

In "De Re Belief," Kaplan comes full circle to accepting certain brute facts of reporting. There is no stopping of a reporter who substitutes his-and-audience in-the-know definite description for a use of an original description by an out-of-the-loop thinker. Thus consider the following variation on Kaplan's own examples. The Saddam-led Iraqi high command is gloating during the March 2003 hostilities, "They (namely, the CIA) believe that Saddam's double is Saddam." Of course, no such assertion was made by the CIA analysts, naïve as they might be. Rather the Iraqi here report the CIA's direct discourse "(We vouch that) the man seen in the video shaking hands with ecstatic supporters is Saddam."

No doubles or deception need induce the free substitution of such content-disparate but referent-sharing descriptions. Back in Paris, speaking of David Kaplan's visit, Pierre may have said in French and using a locally pertinent definite description, "Le professeur avec le costume élégant qui nous a parlait la semaine dernière était amusant." I report all this to you on the Santa Monica Beach, as we witness the splash appearance of that man, in English and with my own pertinent description, "Pierre believes that the surfer in front of us who is just changing into a Hawaiian shirt is funny."

I read the later Kaplan as acknowledging that we simply cannot pull the reporter aside and reprimand him with: "Look here, reporter, you should have read Kaplan's 1977 monograph. Before reporting Pierre's beliefs, you should research carefully how exactly Pierre expressed himself back in Paris. If Pierre used an indexical, then, we would grant you a special dispensation and the freedom to use another co-referring indexical. But if Pierre used a name or a description, you should forget about any such freedom, and make sure you use the very name or definite description that was on Pierre's mind." Any such imagined corrective-legislative constraining of reporters is wishful thinking. The fact of the matter—the matter of reporting people's cognitive lives—is that the reporter is his own master and is free to use whatever expression picks out locally the pertinent referent. And the reporter says something true if the thinker—for example, our Pierre, he as well free to think of the object any way he wishes—applied to that referent, the same property.[18]

---

[17] Kaplan's paper "De Re Belief" was given in the 2001 San Marino conference on the philosophy of David Kaplan and the Pacific American Philosophical Association spring 2003 meetings. Eaker's semantics of reports is given in her 2002 University of California, Los Angeles thesis and in her forthcoming papers,"What We Talk about When We Talk about Beliefs," and "Reporting Attitudes" (in press).

"De Re Belief" encapsulates these findings by the principle I will call the *disunity thesis*:

(DT) Reports are de re, Beliefs are de dicto.

We may have questions about what exactly is meant by describing as "de dicto" the structure of the cognitive fact proper—for example, the belief of Pierre—a question we shall come back to. But this much is clear: whereas reports are object-bound, the disunity thesis insists that the belief proper, as it were the mental episode of thinking *in* Pierre, involves essentially contentful representations. And indeed Kaplan quips "No mentation (for example, belief) without representation." His intent is to mark the content-indispensability in the cognitive fact-proper, in contrast to the content-oblivious character of our reporting practices. Thus, the disunity.[19]

---

[18] In discussing below Russell's yacht problem ("I thought that your yacht was longer than it is"), we shall see that, in like manner, the reporter enjoys substantial freedom in selecting terms to refer to properties (relations) he reports about.

[19] Forms of the disunity thesis appear in the aforementioned works of Evans, Perry, Stalnaker, and van Fraassen. Like remarks apply to Tyler Burge's work, early (for example, "Belief *De Re*," this JOURNAL, LXXIV, 6 (June 1977): 338–62), and late ("Five Theses about De Re States," to appear in Paolo Leonardi and Almog, eds., *Themes from the 2001 San Marino David Kaplan Conference*). I should like it noted that in the contemporary literature (for example, in Burge's work and in Kaplan's early influential work from 1968, "Quantifying In," originally published in *Synthese*, XIX, 1–2 (1968–1969): 178–214), the term "de re state (belief, and so on)" is used to mean what I would characterize as: de dicto state (belief, and so on) involving *nonpurely* qualitative (quasi-demonstrative) representations. The de re/de dicto distinction is in effect used to draw a distinction *among* representing *dicta*: (i) those representing an object by descriptive (qualitative, and so on) means and (ii) those using impure-contextual representations. I made a point of not using—when speaking in my own name—the tainted terminology of "de re." But were I forced to employ it, whether in the context of modal or cognitive reports, I would use "de re" with its original classical intent to mean: altogether representational content-*free*, involving the *object* and nothing but it. Any content-involvement, whether the content is purely qualitative or quasi demonstrative, is, in my sense, is making the cognition not *de*-object but *de*-content.

As will become clear (in the next few sections), in all the works just cited what drives the acceptance of the disunity thesis is the upholding of Frege's *existence* principle, namely, no cognition without representational contents.

In sum, I mentioned above two methodologies of reacting to the quartet's 1970 data, the *semantic reformation* and the *cognitive reformation.* The semantic, admittedly more common, methodology leads to the series of two-dimensional neo-Fregean views we have reviewed, culminating in Kaplan's disunity thesis, a thesis we found already in Frege's *SB* footnote, with its distinction between (i) reporting-by-sheer denotation versus (ii) the essentially content-ful structure of the cognitive fact proper.

I would like us now to turn to the other, *cognitive reformation* diagnosis of the 1970 data, an analysis I associate with Kripke and Donnellan.

IV. REFORMATION METHODOLOGY II—NO CONTENT[20]

A. Simple Cognitions—The Refutation of Frege's Distinctness Conjecture. The insight our analysis is based on is due to Kripke. It comes into full focus in "A Puzzle about Belief," where Kripke presents it as "the clearest objection to the Frege-Russell framework." I quote the key passage in full:

> But the clearest objection, which shows that the others should be given their proper weight, is this: the view under consideration does not in fact account for the phenomena it seeks to explain. As I have said elsewhere, [Kripke refers here to *Naming and Necessity*—JA], individuals who "define" "Cicero" as "The catiline denouncer," "the author of De Fato," and so on, are relatively rare; their prevalence in the philosophical literature is the product of the excessive classical learning of some philosophers. Common men who clearly use "Cicero" as a name of Cicero may be able to give no better answer to "Who was Cicero?" than "a famous roman orator," and they would probably say the same (if anything!) for "Tully" (actually, most people never heard the name "Tully"). Similarly, many people who heard of both Feynman and Gell Mann, would identify each as "a leading contemporary theoretical physicist." Such people do not assign "senses" of the usual type to the names that uniquely identify the referent (even though they use the names with a determinate reference). But to the extent that the indefinite descriptions attached or associated can be called "senses," the "senses" assigned to "Cicero" and "Tully," or to "Feynman" and "Gell Mann" are identical. Yet clearly speakers of this type can ask "Were Cicero and Tully one Roman orator or two different ones?" or "Are Feynman and Gell Mann two different physicists or one?" without knowing the answer to either question by inspecting "senses" alone. Some such speaker might even conjecture, or be under the vague false impression, that, as he would say, "Cicero was bald but Tully was not." The premise of the argument we are considering

---

[20] Throughout this section, I owe thanks to conversations in the mid-eighties with Wettstein. See also note 23 below and the last note of the present piece.

> for the classic position of Frege and Russell—that whenever two codesignative names fail to be interchangeable in the expression of a speaker's beliefs, failure of interchangeability arises from a difference in the "defining" descriptions the speaker associates with these names—is, therefore, false (*op. cit.*, pp. 246–47).[21]

I will call the data mentioned here, (Kripke's) *in the head data*. The in the head data leads to three observations. There are two *negative* points: Frege's *distinctness (of content)-conjecture* is false and, in turn, even Frege's *existence (of content) conjecture* is false. This is accompanied by a third observation: a two-tiered explanation of the *cognitive bond* between a thinker—Pierre—and the object he is thinking about—Cicero.

B. The Refutation of Frege's Distinctness Conjecture. Kripke refutes Frege's distinctness conjecture twice over. There is first a general, *language-free*, structural (I would say, logical) point about any *information processing*; this is followed by a specific application to the case of processing of linguistic items, sentences.

Kripke's structural point about general information processing is this. Let us think of a purely perceptual situation involving no processing of words or sentences. Mary sees through misty clouds a planet; her eye takes a snapshot (think of her downloading a mental "Polaroid-photo" of that snapshot). She next lowers her head to have a sip of water and raises it again. She takes another such snapshot of the sky; then, downloads again her Polaroid-photo. Mary is now considering the two Polaroids, P_1 and P_2 (either inside her mind as "images" or external hard copies in front of her eyes). She is to judge whether P_1 and P_2 are of the same planet. Judge as she will, the judgment is informative for Mary. Why? According to Frege's distinctness conjecture the answer has to be this: only because the two Polaroids had two distinct intrinsic contents (presentations).

According to Kripke, this is simply not so. The informativeness may rest precisely in the *sameness* of the perceptual information (presentation). The point is that Mary is given twice: a planet looking thusly (she is not given it in words; I give us her "image" by a verbal description). Her qualitatively given "data base"—planet looking thusly—cannot settle whether one and only one planet has been presented. This data base is too weak—insufficient—to decide the truth (falsehood) of the identity judgment. The question whether the two Polaroids are of the same planet requires going beyond the intrinsic content of each.[22]

---

[21] A common retort is that I must have in the head associated with 'Cicero' the distinguishing feature "designates someone called (by me) 'Cicero'" (and likewise for 'Tully' namely "designates someone called (by me) 'Tully'"). As against this, Kripke annotates (note 17 in the original, as well as note 12) the present quoted passage with the observation that "called by me 'Cicero'" is not really more definite than "Roman orator," for many are so called by me (for example, a certain twentieth century famous German spy and my neighbor's dog). I would add that in thinking of that Roman man, I need not have in mind any linguistic thought about names, designation, being-called, and "by-myself." I just think of *that man*, Cicero.

The identity judgment's informativeness does not consist in some "informative propositional content," a supposedly *intrinsic* feature of some proposition (thought). What holds the key is rather an extrinsic *relational* fact, articulated from what Donnellan calls aptly the *omniscient observer of history point of view*: whether the data base D of our perceptual agent Mary about the perceived planet—the information she does have in her head—can settle whether P_1 and P_2 are of the same planet. But D is incomplete—leaves open—the status of this question. It is this relational fact spotted by the omniscient observer—the *failure* of her specific data base D to settle the identity—that makes it informative for *her*. In sum, Kripke inspires us to submit:[23]

> (K1) *Informativeness as data base incompleteness*: the informativeness rests in the agent's internal data base ("theory") incompleteness—it underdetermines the truth value of the identity judgment; both the identity and its negation are "live options" for the agent.

---

[22] I am not here *asserting* that such a notion of "purely qualitative" or "intrinsic content" of the picture ultimately makes sense (I am skeptical about this). I rather use the notion to make a point about its *limits*. Also, let it not be assumed that the intrinsic content has a time index, for example, "planet as seen thusly at *t*," an index separating the two pictures. That the picture was taken at *t* is true. But what flashes before the mind does not include such reference (by the time Mary contemplates the pictures, she may well have forgotten when they were taken and in what order).

[23] Kripke's quote makes the point about simple cognitions in passing (his main focus is indirect discourse reports of simple cognitions): I have come to appreciate the point in discussions in the mid-eighties with Wettstein, who has since amplified on it in a series of papers and in his recent book *The Magic Prism* (New York: Oxford, 2004). See also a forthcoming symposium on the book in *Philosophical Books*. In my own contribution to it, "Direct Reference and Significant Cognitions: Any Paradoxes?" I enlarge on Wettstein's insights.

Let me lay to rest the following worries raised in discussion of (K1) by various neo-Fregeans (perhaps following Frege's own hints in his pre-*SB* 1879 *Begriffschrift*) and put most forcefully to me by Kaplan.

Such audiences tried to reconstruct my point in (K1) as asserting that the name 'Cicero' does have a content after all, namely, *the referent of 'Cicero'*; what I here wonder about is the *propositional content* of "The referent of 'Cicero' is the referent of 'Tully'" (shades of Stalnaker's two-dimensionalism). As against this (see also Kripke's remarks from note 20 above), I will say: (i) *The referent of 'Cicero'* is not a uniquely identifying content, (ii) ordinary speakers, for example, myself, do not reflect on such theoretical-semantical notions when they *use* the word 'Cicero' (I am closer to having Roman oratory in mind when using 'Cicero' than any reference of names issue), and (iii) logically speaking, the explanation given in (K1) involves no content-assignment to any

(K1) concerns the source of the informativeness of identity judgments, any allusion to language aside—the case dealt with was perceptual and as such accessible to nonlinguistic agents, for example, babies and dogs. We may now move to Kripke's application of (K1) to specific cases where the information processing is carried on through *linguistic* media, for example, using proper (common) names. Kripke now observes that most of us learn names such as 'Cicero' and 'Tully' while associating with each nothing more than an indefinite piece of information coded by the indefinite noun phrase "A Roman orator." This incomplete information—the internal data base of the linguistic agent about that man—underdetermines, for that agent, the truth value of "Cicero=Tully."

C. Kripke versus Frege: Truth and Cognitive Significance. We must take notice of the fundamental difference in the order of explanation offered by Frege and Kripke. For Frege, *cognitive significance precedes truth.* First, the names 'Cicero' and 'Tully' are associated with their respective "senses," say [The *F*] and [The *G*]; a proposition (thought) is generated < [The *F*] [is] [The *G*] >. This proposition is *already* different from the one generated by "Cicero=Cicero" namely, <[The *F*] [is] [The *F*]>. Indeed, quite independently of any *evaluation*—checking for the denotations of the two senses and the truth value of the propositions—namely, by sheer analysis of the *intrinsic content* of the two propositions, we already see the cognitive difference immediately after their generation and prior to any evaluation, any assessment of truth value.

With Kripke, we get reversed priorities: *truth precedes cognitive significance.* The truth of the public language sentence "Cicero=Tully" is fixed by the vicissitudes of history *prior* to the question of what information any cognizer, for example, Pierre, acquires in order to "have" the names. Prior to and *independently* of this information—the historical mechanism binding the names back to this Roman has nothing to do with the contentful information Pierre *subsequently*

---

name. It involves a *logical-computational* relation between a *set of sentences* (what I have in the head, my "Cicero" theory and my "Tully" theory) and an object language target *sentence*, "Cicero=Tully." The informativeness is thus relative to two parameters: (i) the given data base of the agent, and (ii) his/her reasoning engine (the logic-calculations feasibly available to this agent who is not logically omniscient). If relative to these two parameters, only one given truth value is assignable to the target sentence, it is not informative; if its truth value is not so determinable (both values are consistently assignable), the target sentence is informative. This relational notion of informativeness (truth value calculation vis-à-vis the two parameters of (i) the agent's data base and (ii) the agent's reasoning engine) does not involve, anymore than the classical formal notion of deducibility (provability, and so on) does, any content-assignment.

comes to associate with the names. This "subsequently" is critical to Kripke's explanation: the truth (falsehood) of the sentence "Cicero=Tully" (for that matter, see below, "Paderewski=Paderewski") is already fixed, by external history, before any cognizer comes in. Good luck to the cognizer: the scant information he picks while learning the names ("A Roman orator" and again "A Roman orator") will have to settle, "from the inside," what the history of the world has already fixed on the outside.

This style of explanation of informativeness "from the omniscient observer's point of view" has the following additional consequences (K2)–(K5):

(K2) Informativeness is relative and relational.

The notion of the informativeness of "*a*=*b*" is not *absolute*, as in Frege, but *relative* to the agent described from above. In the Fregean content solution, "*a*=*b*" is assigned one fixed propositional content—either that content is informative or not, never mind who grasps the content. The informativeness is inherent to that propositional content.

In the "from the omniscient observer point of view" explanation, "Cicero=Tully" may be informative for you, not for me; or for me now, but not later. The reason: the informativeness is a relational feature connecting a given data base to "Cicero=Tully." In terms of (in)completeness of the agent's theory, we can see that the status of "Cicero=Tully" is relative to the richness of the theory from inside which we try to settle it.

(K3) "*a*=*a*" may well be informative.

On the present approach, English (ordinary language) "*a*=*a*" cases may be as informative as "*a*=*b*." Thus "Paderewski=Paderewski." The competent user's data base underdetermines the truth of the identity sentence. This is not because the two occurrences express different contents. It is because all one has in the database for the one 'Paderewski' is "A famous Pole" and for the (other) 'Paderewski', again merely "A famous Pole." This much does not allow to derive the truth of "Paderewski=Paderewski"—the information is consistent with the falsehood of this sentence.[24]

(K4) Frege's own dilemma—problems with the ideal "The *F* is the *G*" cases.

---

[24] I, for one, never understood how Frege can ask us a question (regarding apriority or cognitive significance, and so on) about the purely *schematic* "*a*=*a*" and how all of us

Frege would like us to reduce the explanation of the informativeness of "Cicero=Tully" to that of the descriptive proxy "The author of De Fato is the denouncer of the Catiline." I have emphasized the inapplicability of the strategy by pointing out that for many such names, we have no descriptive proxies in the head. But I would like now to make a stronger claim: it is presupposed throughout this discussion that Frege's content program does explain well the informativeness of "The *F* is the *G*" sentences. Indeed, Frege's explanation by content is often presented as the only possible explanation. I would like to submit that this is quite wrong. Frege's explanation of informativeness by content fails even for "The *F* is the *G*" cases.

Thus, in all, I lead into the following odd conclusion—where we all find an intuitive *pre*-theoretic puzzle—the so-called Frege puzzle about how could "Cicero=Tully" be informative—I find no puzzle; but where we all find an elegant solution—Frege's *theoretical* account of the informativeness of descriptive cases like "The evening star is the morning star"—I actually see a puzzle (for Fregeans only), what I will call *Frege's own dilemma.*

Consider Frege's claim "The evening star is the morning star" (and keep in the back of your mind its twin "The evening star is the evening star"). Frege needs to guarantee that his account of the informativeness of the claim respects at once two desiderata:

(i) *The referential condition*—(ia) "the evening star" and "the morning star" are each used to *refer*, (ib) indeed to *co-refer*—to some particular *astronomical object* and (ic) thus, "The evening star is the morning star" is an *objectual-astronomical identity.*

(ii) *The informativeness-by-content condition*—the one identity's content is informative ("The evening star is the morning star"), the other ("The evening star is the evening star") is not.

I submit that Frege cannot satisfy simultaneously both desiderata. Either he primes informativeness-by-content and has to sacrifice the referential condition or he sustains the referential condition and the informativeness-by-content has got to give. This either/or I call *Frege's own dilemma.*

---

respond to him automatically without protesting that the question is not well posed. For me, "$a=a$" is a sheer *un-interpreted form.* We can say about it that it is *formally* provable in (a system of first order) formal logic. But unless interpreted (or unless we plug in an *instance* from English having the same *visible* syntactic form, for example, "Cicero=Cicero"), I cannot see how to answer the question whether we know the *truth* a priori or whether the *truth* is informative. The purely schematic "$a=a$" is not something that can be said to be true (false).

We can bring out the first horn of the dilemma by attending to the case of "The evening star is the evening star." Suppose we say that by pure inspection of this intrinsic content (sense), we can "see" the claim is un-informative. We thus respect the informativeness-by-content condition. But do we still respect the referential condition?

Frege's account of informativeness-by-content is totally prior to and independent of the state of the heavens, in particular of the existence of Venus and of *its* features. For—and this is now the key fact—Venus is simply *not a star.* If God (i) had not created Venus to start with and (ii) whether He made it a star or not, Frege's account of informativeness already applies, regardless altogether of the worldly Venus and its ways-of-being. Frege settles the matter of informativeness—of "The evening star is the morning star" as contrasted with "The evening star is the evening star"—by sheer content analysis, *before* any evaluation of the content for (co-) *denotation.*

The procedure we just followed—x-ray analysis of intrinsic contents—applies just as well to the difference, and Fregeans automatically insist there is one, in the denotation-*less* cases of "The planet perturbing Mercury is the planet perturbing Mercury" (un-informative) and "The planet perturbing Mercury is the planet posited by Leverrier and Adams in the mid 19th century" (informative)."

Through and through, Frege's explanation of informativeness-by-content flouts then the referential condition on these singular expressions. Frege could give this very explanation by eliminating the descriptions from the language (as urged by Quine) and use instead sheer predicates, for example, point to the difference in content between the purely general-quantificational "Everything is such that: it is evening star iff it is evening star" versus "Everything is such that: it is evening star iff it is morning star." Following this route, Frege explains the informativeness of *something*: the informativeness of one original nonreferential sentence, the uninformativeness of an *extensional* predicate-transform thereof. He does not explain the informativeness while making the singular expressions in the sentence (co)-*refer* to Venus.[25]

---

[25] I should like it noted that were Frege to follow the letter of his own theory, it predicts that "The evening star" and "The morning star" may well (and obviously contrary to our use in English) denote a star—call it Gotl—at the other end of the galaxy that we never set our eyes on (and had no other light-signal from), simply because, as cosmic luck would have it, Gotl happens to satisfy strictly (both) predicates (and hence be the jure *bedeutung*). If Frege would say at this juncture that Gotl is the pertinent item, he surely is not explaining the (his own!) idea that these two English descriptions are *ways of giving us* modes of presentation of...Venus. He would not be explaining the *cognitive* significance this identification has for us.

The other horn of Frege s own dilemma is forced on us by insisting on the referential character of our descriptions (and the astronomical-objectual identity induced). The object-level English sentence "The evening star is the morning star," as we *use* it, is true indeed; but it is *not* true *because* the two contents (senses) denote a single object, namely, Venus. The two singular expressions make the sentence actually true because both *refer* to Venus, in spite of Venus's *not* being the *bedeutung* (*denotation*)—not satisfying the *de jure* content(sense) of the descriptions. The two descriptions *refer* to Venus because it is that object, Venus, that made us, by an appropriate historical process, generate these (false) descriptions for (ways of being given of)—*that thing*.

On this last account, we respect the referential condition: the two descriptions co-refer; they co-refer to the historically right item, *Venus*; and yes, the claim "The evening star is the morning star" is a true astronomical-objectual identity. The explanation of the informativeness of "The evening star is the morning star" must now respect these three facts and operate *while washing away* the contents of the descriptions. For recall that the referential condition has been met precisely by ignoring (and in spite of) the content of the descriptions. The content did not fix for the user of the descriptions with his object of thought—Venus; the content was the users's subsequent attempt to communicate which object he is already thinking of. I suggest that, as in our above explanation for "Cicero=Tully," the claim is informative for many thinkers because their astronomical data base is consistent with the falsehood of "The evening star is the morning star." On the other hand, "The evening star is the evening star" also respects the referential condition: the descriptions refer, indeed co-refer, and they induce an objectual-astronomical identity (regarding Venus). Yet, the claim is not informative: it is not consistent with my data base that the "The evening star is the evening star" is false.

In a nutshell, for Frege, it is either informativeness by content but not of an objectual (astronomical) identity or else it is an astronomical-objectual identity all right but we explain its informativeness by a content-free procedure.[26]

---

[26] The case of "The evening star" and "The morning star" (it is Frege's own) is of course not special. Consider a scenario where even Kripke claimed (wrongly in my view) a priori knowledge is available (via the use of a reference fixing description), the case of Jack the ripper. Suppose the detective in charge of the case reports: "The man who raped all these gals is the man who strangled all these gals." Now, assume the murderer is as a matter of fact a woman, Jill, disguising herself as a man, and faking the semen and strangling evidence to mislead the police (in fact, she calmly poisoned the gals way back). The objectual-criminal identity is very informative (for the police as it is for us).

D. The Refutation of Frege's Existence Conjecture. Finally, I turn to our last, perhaps most basic, feature of the "from above" explanation, (K5). (K5) has its source in a Donnellan-Kripke very general observation about information processing about objects, be it in sensory perception, nonperceptual thought, or linguistic reference. I will call this thesis the *priority of the cognitive bond*:

> (K5) The determination of the object of cognition—for example, Cicero—is prior to and independent of an association of some predicative content (for example, "a Roman orator," "the author of De Fato") by the cognizer.

The best illustration of the priority of the cognitive bond thesis I can think of is due to Donnellan. You and I are in a party facing a man, John, with a martini glass (unbeknown to us, his glass is filled with plain water). I apply to him, in my thought, the predicate '...is tipsy'. As luck would have it, there is, hidden from us back in the kitchen, one unique man in this party, George, drinking a martini. My plan is to externalize ("communicate") to you my thought that John is tipsy. And to direct you to my object of thought—the *already fixed* object of thought—I say "The man drinking a martini is tipsy."[27]

To reiterate: when the question "Tipsy who?" comes up, I am already fixed on my object of thought; I only need to direct *you* to that-fixed-object. In this respect, whether the information I provide when asked "Who is Cicero?" is definite or not is rather anecdotal. The critical question is what the information *does*, where it comes in. Of course, Kripke's clever use of cases where I have only *in*definite information ("a Roman orator") is dialectically powerful because it is clear that this indefinite information cannot fix the cognitive bond. But, in truth, even if I have a definite description to provide, the role of this information is the same—communication enhancement.

But it is not explained by content: the predicative contents are false of the criminal Jill. In spite of this, what makes the identity claim true (and informatively true!) is none other than this single woman co-referred to, Jill.

[27] There is here an independent question in the semantics of definite descriptions in ordinary English—Is the sole semantic denotation ("semantic referent," says Kripke) of the description the kitchen-man, George, or is there an ambiguity in the language and is there, in addition, a *convention* based ("referential") reading of the description making John *its* semantic referent?

I believe (against Kripke) the latter and have argued about this semantic question in "The Subject Verb Object Class II," *Philosophical Perspectives*, XII (1997): 77–104, and yet in more detail recently in "The Proper Form of Semantics." This semantic issue about the conventional rules governing such a phrase in English is not our current topic.

Viewed in this way, as a subsequent *communication helper* rather than as the *cognitive fixer* (to use Wettstein's pregnant phrase), we can see why Frege's account of the predicative information is false. To remind us, the idea of Frege's existence conjecture is that there is no cognition of an object—for example, Cicero, without a representation, for Frege, without a uniquely identifying predicative content. In reading the conjecture as false, we must discern various readings. To be sure, there are readings of the conjecture that make it true. For example, it is very correct to say "no thought about Cicero without (at one point or other) the bouncing of photons from Cicero" and "no thought (of mine) of Cicero without the existence of my brain," where I simply record metaphysically necessary conditions for the existence in the world of an about-Cicero-thinking (by me). In this sense, it may well be true that in every possible situation in which I am thinking (seeing, referring) to Cicero, there is indeed not only photons that bounced from Cicero's surface and physical particles that penetrated (later) my sensory channels but there must also be my living brain; after all, we surely want to say "no mentation (of Cicero) without a living brain." If the mentation is one of visual perception, there may have to be on top of the foregoing, an *image* in my mind. If the case involves linguistic reference, there may have to be in my mind a *word*. And finally, still in this sense of necessary conditions for the cognitive state, we may well say that if the mentation is an act of thinking, there may have to be in my mind some representational medium, better even, some *trace* (be it an image, word, and so on) *of* Cicero, whenever I think of him.

Without these background facts—the existence of photons, brains, images, words, and so on, there would be no vision of, reference to, or thought about Cicero. We articulate here necessary conditions for the occurrence in the world of a certain (mental) act, just as we could articulate necessary conditions for the occurrence in the world of other acts of mine, for example, "no kissing without touching" or "no begetting of my son without producing a sperm to fertilize an egg." There are necessities of kissing and begetting and there are necessities of seeing, referring, and thinking.

In this sense, "no mentation without representation (images, words)" may well be true and necessarily so. But this is not the substantial sense targeted by both Frege and Kaplan. The substantial sense does not merely record the existence of a representation-trace as a necessary condition for the existence of a thinking-act. Rather, on the substantial reading of the existence conjecture, we are interested in the existence of a representational content which *determines*—perhaps together with the *context of thinking*—the object of the thinking relation. Indeed, Frege (and Kaplan?) are wedded to yet a stronger claim: the existence of a content—even if the two provide us with a different *kind* of such content—that (i) determines the object of my thinking, for example, Cicero, and (ii) that I the thinker can *provide* when asked "And who for you is Cicero?" The content is both (i) metaphysically *determinative* of the object-of-the thinking relation and (ii) epistemically *available* to the subject of the relation (and it is only because it is so thinker-available that it is going to explain, for that thinker, the cognitive significance of "Cicero=Tully").

It is these substantial readings of the existence conjecture that I see as refuted by Kripke and Donnellan's observations about (i) what *does* bind my thinking to Cicero, and (ii) the merely posterior (to cognition-fixing) communication-helping role of the predicative content I associate.

V. CONTENT-FREE EXPLANATION OF REPORTING-COGNITIONS

A. Multi-thinker Puzzles. The well-known puzzles generated by reporting cognitions—for example, those involving Kripke's Pierre or Quine's Ralph—specify a case in which only *one* thinker is concerned. We are then pondering a plurality (for example, two) of the cognitive states of that single mind. Goaded by Frege's *SB* footnote insight about indirect discourse, I would like us to consider generalizations of the foregoing stories in which we have more than one thinker—a plurality of agents.

The Generalized Pierre Puzzle. Pierre says "Londres est Jolie" and answers "What is Londres?" with "La capitale d'Angleterre"; Francesca says "Londra e bella" and responds to "Cosa e Londra?" with "La citta dove si trova Big Ben." I report it all to you in English with "Pierre believes that London is pretty and Francesca also believes that it is pretty" (or: "Pierre and Francesca believe (both) that London is pretty"). They seem to agree in their attitude but how could this be if their thought contents are different? In a variant case meant to produce a contradiction, Francesca, still backing 'Londra' with "La citta dove si trova Big Ben," says "Londra non e bella." I report with "Pierre believes that London is pretty and Francesca believes that it is not," a true report read by many to suggest that Pierre and Francesca contradict one another. But how could they contradict one another if their thought contents are not contradictory?

The Generalized Ralph Case (due originally to Quine). Ralph says "the man on the beach is a spy" and Bert says "The man with the sophisticated binoculars is a spy." Seeing the man, in a brown hat, cruising before us, I report "Ralph believes that the man in the brown hat is a spy and Bert also believes that he is a spy." The truth of the report suggests the same attitude on the part of Ralph and Bert. But how could they co-think if their thought contents are different? In the variant contradiction case, Bert says "The man with the binoculars is no spy." I report later "Ralph believes the man with the brown hat is a spy and Bert believes he is not." I seem to assert that Ralph and Bert contradict each other. But how could they if their thought contents do not?

The Generalized Empty Name Case (due originally to Donnellan). François says "Père Noel est genereux," identifying the man as "the guy who comes down from the north pole with presents"; Paolo says "Baba Natale e generoso" identifying him, in English, as "the guy who comes in a reindeer pulled cart from Finnmark"; I report them both in English "François believes that Father Xmas is generous and Paolo also believes he is." But how could they share their attitude if their thought contents differ? In the variant case, they seem to contradict each other, after Paolo says "Babo Natale non e generoso." I report this with "François believes that Father Xmas is generous and Paolo believes that he is not." But how could we have two contradicting fellows if their thought contents are not contradictories?

B. Frege's Normal Form—Coordinative Reports. With his remarkably "musical ear," Frege spotted (perhaps in spite of himself) in his *SB* footnote the simplest and at the same time most fundamental form of all indirect discourse reporting cases. I would like to call this the *normal form* of indirect discourse reporting. What Frege noticed is that our account of indirect discourse cannot be reduced to very special limit cases. For example, consider the case of a single mind at a given time that is both playing the thinker and reporter roles—I think to myself "I am happy" and I automatically report "I believe that I am happy." This special set up cannot be assumed in general.

In the general case, the reporter and the thinker will drift apart. Thus even if we have solely one thinker to report on, the reporter may have to shift: indexicals, the language used, the time (of the *report*, as opposed to the time of the thinking). And *one*-thinker-at-*one*-time to report on is still too special a case. For in the yet more general case, we have a *series* of thinkers (or mind states of a single person through time). This is what Frege's footnote agonizes about: there is your association with 'Aristotle', there is my different association, and there is now a third party, a reporter, that has to put it all together *in a common currency language* and report that both of us believe that Aristotle was clever.

Frege's reporting normal form focuses us on both distinct syntactic and semantic markers of such common currency indirect discourse reports, what I shall dub—*coordinative attitude reporting*. Syntactically, we have encountered forms like "John and George (both, each) believe that Aristotle was clever," "John believes that Aristotle was clever and so does George," and most explicitly (with the pronoun marking the coordination) "John believes that Aristotle was clever and George (also) believes that he was clever."

As we analyze the contradiction ascribing puzzles, the coordinative anaphoric structure will manifest itself. For example, in Kripke's Pierre story, we find the report "Pierre believes that London is pretty and he believes that *it* is not pretty"; in Quine's case, we have "Ralph believes that the man on the beach is a spy and believes that *he* is not a spy"; and in Donnellan's case, we get again, "François believes that father Xmas is generous and François believes that he is not generous"; and so on. The reporter has introduced his—reporter's—coordinative subject-*noun* (that may well be unknown to or unused by the thinker(s) reported). The *pro*-nouns that follow are again the *reporter's* markers, indicating he is now wiring a host of cognitive states (thinkers) onto his elected coordinative subject.

C. Coordinative Reports II: Truth Conditions. What are the truth conditions of the coordinative anaphoric reports?

Let us focus first, as Frege does in his footnote, on reports using denoting names and descriptions (a natural generalization will be called below for reports with empty names (descriptions)). Suppose Frege reports "Aquinas believed that Aristotle was clever and Descartes also believed that he was clever." We may suppose Aquinas identified Aristotle as "the philosopher whose metaphysics I am about to explain to you" and Descartes identified him as "the philosopher whose metaphysics I am about to refute for you." We may well take Frege to identify Aristotle as "the student of Plato and the teacher of Alexander." What makes Frege's report true?

Aquinas used some term (name, description) '*A*' to say "*A* was clever" in his original language (Latin, Italian?); and (unaware of Aquinas's assertion) Descartes used some term '*B*' to say, in French, "*B* was clever." I will make Frege use English. For the report in Frege's mouth to be true, our three protagonists need to be historically *co-wired*: the origin of the reporter's-Frege's—'Aristotle' has the very denotation that is the historical source of Aquinas's cognitive state (of his thinking "*A* was clever") and independently, of Descartes's state ("independently" for there may have been no communication between Aquinas and Descartes).

It is important to take notice of the following fact of reportage: the reporter, Frege, does not record in his report *how* either Aquinas and Descartes expressed their states; nor does he record how he himself identifies Aristotle. The only thing that matters is that, as a matter of (omniscient observer ascertainable) historical fact, it is this one object—Aristotle—that induced their cognitive states.

Since we still have the discussion of the disunity thesis ("reports are de re, beliefs are de dicto") in the back of our minds, I would like us to notice that there is here no gap between our indirect discourse reporting—markedly de-*denotation*—and the Aquinas/Descartes cognitive states reported—allegedly each essentially de dicto (de-content). To the contrary: our reports are *reflective* of the nature of the cognitive states. The cognitive states of Aquinas and Descartes are individuated in terms of the object the two gentlemen are related to and what property they apply to it, and regardless of how they might help us later identify—communicate to us—which item was their cognitive-object. And so, Frege's report is anything but *abstractive* of the cognitive situation; it is rather made true *by* the fact that Aquinas and Descartes are in the *same cognitive state*—co-thinking that Aristotle was clever.

D. Coordinative Reports III: Pierre, Peter, Ralph, and François. We can now see how our various puzzles are instances of the Frege normal form. Thus consider our report on London-bound Pierre. We have a thinker's use of 'Londres' in French; next, a thinker's use of 'London' in English (in my own telling of Kripke's story, I always make sure that Pierre's second use is in yet a different language, for example, Italian, so that the reporter's ultimate language is distinct from both of the thinker(s)' languages). The first thing to notice is that the reporter's sentence:

> Pierre believes that London is pretty and Pierre believes that London is not pretty

may be put in the form of a coordinative anaphoric chain:

> Pierre believes that London is pretty and Pierre believes that it is not pretty.

The 'it' is the reporter's anaphor. By using this form of report, the reporter keeps track of the subject of discourse that he, the reporter, introduced via 'London'. And the report is true. For there exist expressions 'A' and 'B', co-denotative with the reporter's 'London', such that Pierre assented to "A is pretty" and to "B is not pretty." Nowhere does our report suggest that Pierre is contradicting himself, as if he assented to "A is pretty-and-not." This last kind of assent would have been a threat to his rationality. But Pierre is involved in no assent of the sort.

In the present framework, the Paderewski-story follows the same path as the Pierre-London tale. I report:

> Peter believes that Paderewski is musical and that he is not musical.

Nowhere do we ascribe to the inner life of the thinker Peter any conjunction of thoughts, using a logically univocal name '*A*', of the form "*A* is musical and not musical." Perhaps all Peter said was first "The Polish musician I heard a concert of is musical," then "A Polish high official I met at the airport is not musical." The reporter's description is not busy providing the relevant *evidence* for the truth of his own report; it provides the key *object*—using a favorite expression of the reporter—and the cognitive *relations* Peter bears to that object.[28]

Our method resolves in a similar manner Kaplan's suspender of judgment case. The reporter says:

> Pierre believes that London is true and does not believe that London is pretty.

The report is true (without making *us* irrational) because (i) there is an episode in which Pierre assents to a claim of the form "*A* is pretty" (with '*A*' London-designating) and (ii) there is a London-designating term '*B*' such that Pierre *does not* assent (suspends judgment) when asked whether "*B* is pretty." This is enough to make our report true.[29]

Finally, we resolve Quine's Ralph case. Our report reads:

> Ralph believes the man in the brown hat is a spy and Ralph believes that the man in the brown hat (he) is not a spy.

Quine hears this to be (i) true and yet (ii) not imputing irrationality to Ralph. I agree about the truth assessment (as about the rationality): there exists a denotation-sharing term, namely, "The man on the beach" for which Ralph assented to "the man on the beach is a spy" and a denotation sharing term, namely, "the mayor," for which Ralph assents to "The mayor is not a spy." As noted by Quine (and anticipated by Frege's footnote insight regarding *denotation-sharing* being sufficient), the key to the solution is not the use of names (for example, 'Ortcutt') as opposed to descriptions. The Frege-Quine insight cuts across the name/description divide and focuses rather on the thinker/reporter divide. Frege's footnote and Quine insist that: the report is made by the reporter. He gets things right as long as he, with *his* terms—names or descriptions—keeps the same denotation as those of the thinker(s).

---

[28] For more in this vein on Peter's rationality, see the remarks on the very idea of rationality and internal-form in the concluding section below.

[29] From the present perspective, Kaplan's solution in the 1968 "Quantifying In" (insisting that *our* language must quantify over modes of presentation) runs together (i) the object-language of the report (English!) and (ii) the meta-language in which the omniscient observer specifies *what makes* an object-language (indirect discourse) report *true*. The conditions (i)-(ii) just cited are the omniscient observer's facts making true my marketplace indirect discourse report. The omniscient observer provision of truth-grounds is not (a translation of) *what I say* in using indirect discourse English.

The foregoing leaves us with one final case, those involving empty terms in the report, in particular the above-mentioned Donnellan puzzles. We cannot say here that the reporter's term must share denotation with the one employed in the thinker's cognitive state, for no denotation is to be had. But the solution is at hand in our *pre-*reporting notion of cognitive state—When do we take two individuals to be *co-thinking about* the same thing?

Let us reconsider Donnellan's François case. François's thought (and speech) concerns a "nonexisting" individual. At the end of the story, we report François as follows:

> François believes that Father Xmas is generous and François believes that he is not generous.

The critical question about the unification of our three cognitive states—two episodes of François, the third is the reporter's use—was and remains their historical co-wiring from the omniscient observer's vantage point: Is the source phenomenon behind my use of "Father Xmas" the very source of the two cognitive states of François? The one difference is of course in the nature of the source phenomenon. When we spoke of Aristotle, we picked out (what was) a real object; in speaking of "Father Xmas" we do not. The difference is critical for some issues, not for others.

It is certainly important for a variety of existence-sensitive predication. For example, may we existentially generalize over the report and say "So, there really exists this item, Father Xmas, that is believed by François..."? No, we cannot. Similarly, the reporter can add truly to his long report about François, his own aside "...but he—Father Xmas—never really existed." He would not be able to so add in the case of Aristotle.

On the other hand, there is a cluster of issues for which the difference does not matter. Chief among them is the question: What makes the trio of protagonists "think about the same thing"? Some common content inside their heads—unification *from inside the head*—or a connected history—unification *from the omniscient observer's point of view*? Our answer is that what matters is the *connection* between the thinkers—and between them and the reporter—not whether at the very origin of the connecting wire lies a real object.

VI. CONCLUDING REMARKS

I would like to summarize our findings by running through the main themes of the present work: the reporter-bound character of indirect discourse speech, the unification of language and thought, the rationality of our puzzled agents, and, finally, what got us going, the content dilemma and its impact on Frege's unification project.

VI.1. The Reporter-bound Character of Indirect Discourse Speech. The foregoing discussion focused on puzzles involving singular terms in the embedded clause. But the fact that the report sentence is the *reporter's* sentence, using his own vocabulary, has consequences that reach far beyond substitutions of singular terms.

Consider, for example, universal instantiation. Our departmental secretary Betty may say "Every professor late with his grades will pay dearly." Late professor that you are, I am right to report to you "Betty said that you will pay dearly" (and pay you will!). Indeed, you can confront Betty later in the office and say to her: "I hear that you said that I will pay dearly."

Consider next my conjunctive report (made to Pierre) "You said that London is pretty and not pretty." When the omniscient observer shows Pierre the record of his assent patterns, he will have to concede my report, even though he never said, in the same breath, "*A* is pretty and not pretty." It is thus that Ralph, a movie reviewer, uses a sequence of content-variant but denotation invariant terms for Nicole Kidman, saying in his long review piece "*A* is from Australia" and "*B* has been marvelous in *The Human Stain*" and "*C* is surprisingly authentic in *Cold Mountain*" and.... All this I report truly to you (not ever standing the chance to recall which term went originally with which predicate) with "Ralph said that Kidman was from Australia, marvelous in *The Human Stain*, and surprisingly authentic in *Cold Mountain*."

A third and final example of the inevitability of such transformations was spotted early on by Russell (who, alas, tried to suppress it by regimenting it into the classical content program, using his theory of scope distinctions). Sir Alf owns a yacht, Godiva. You, John, have not seen it but you sure have heard about it. You are impressed. You say to me: "Godiva is huge"; or perhaps you say, "It can take 200 people for a wild party"; or if you are engine minded "It has the petrol tank of an ocean liner." In my next meeting with Sir Alf, I report:

(A) John believes that your yacht (Godiva) is larger than it is.

Of course, if we take the report to display the form (and content) of the thinker's thought, this would make John irrational. This will not do. Though John is over impressed, his logic is fine. Russell rushed to the rescue with heavy equipment: (i) his theory of scope distinctions, and (ii) quantification over lengths (it is not often noticed, because Russell himself passes this by in a sly way, intimating his scope theory *by itself* is enough for this case). Russell claims that the true reading we hear makes me *say* the following:

> (B) There is a certain length, call it a Russell, it is Godiva's actual length, and John believes that Godiva's length is bigger than a Russell.[30]

What should we make of (B)? First, we should notice that Russell himself, for all his protection of the thinker-oriented theory of indirect discourse, indulges here, discreetly, in a shift to a reporter-oriented analysis. For the comparative part of the sentence (B) is not in John's, the thinker's, mind. (B) is meant to be the true content of what the *reporter* says to us from the outside—from his reporter's perspective—about the episode, attributing only codedly a belief in a certain Godiva-length to the thinker. This brings us to the second and more critical point. I believe Russell misses the crux of the phenomenon he "smelled." It is this. The reporter—myself—does not and need not talk about such abstracta as lengths; in turn, the thinker may not have engaged in any comparative thought whatsoever (indeed, as I set things out, he just said: "Godiva is huge" or some such impressive noncomparative property). And other things being equal, I submit that we should take my indirect discourse report at face value—if I do not *mention* lengths, I am not talking about lengths. Any length-talk notwithstanding, my report is true: the source of my—reporter's—use of 'Godiva' is an entity believed by John to be longer than it is. How so? For *evidence* I would cite the fact that John said of that object, using some term '*T*', "*T* is huge (can accommodate 200 people, and so on)." Now, using my own term 'Godiva', I would like to record the *fact* that: "Godiva is not huge but just standard size (can accommodate only thirty people, and so on)." So, this entity that is the source of my use—Godiva—is indeed believed—by a very fine rational man—to be larger than it is. We witness here the reporter's freedom to alter not just the *terms* used by the thinker for (i) the subject and for (ii) the predicate; the reporter alters the *form* of the thinker's thought (which was not comparative).

In an oral presentation of this material, Kaplan engineered a cleverer example of this phenomenon by combining it with the aforementioned fact that reporters often coordinate between a multiplicity of disparate thinkers. In his tale, we add to our over-impressed John, another yacht-observer, Fred, who tells me, in separate circumstances (and in total ignorance of John) "Godiva can take a hundred people on board." I can now report correctly:

---

[30] This formulation of the reading was made to me by Kaplan.

(C) John believes that Godiva is twice as large as Fred does.

Each of John and Fred uttered and thus expressed a noncomparative belief. I, the coordinative reporter, make out a comparative claim out of theirs. What it compares now is not John's belief with Godiva's actual state, it compares the beliefs to one another. None of this—in form and content—can be found in their original speeches-thinkings. Nonetheless, my report is true, not just "approximately" or "pragmatically" and so on, but as strictly true as a report could ever be: John believes that this yacht can take 200 people on board; Fred believes it can take only a hundred; Arithmetic would vouchsafe the truth of my claim that John believes that Godiva is twice as large as Fred does.[31]

VI.2. The Very Idea of Rational Agent (Action). I insisted through and through that the empirically-unlucky Pierre and Peter are perfectly rational agents. But how could this be if the very *form* of what (Paderewski-bound) Peter accepts is "*a* is *F* and *a* is not *F*," a contradictory form? We ourselves, in indirect speech, seem to ascribe to Peter and to London-bound Pierre a contradictory form when we use the form "Pierre (Peter) believes that *a* is *F* and that *a* is not *F*." The problem that lurks here in the wings is that, after all, there are genuinely irrational agents in the world. There must be *some* genuinely irrational agent, let us call him Max, who accepts, just like Peter, the sentence "*a* is *F* and *a* is not *F*"; what is more, Max is reported-true by the same kind of report describing Peter and Pierre, namely, "Max believes that *a* is *F* and that *a* is not *F*." But if so, the *form* of what we accept (report) would not settle anymore whether the agents in question are (ir-) rational. How are we going to tell apart the sane Pierre/Peter pair from the insane Max?

---

[31] As the reporter departs from both the original form and content of the thinkers' episodes, we might wonder, whether any constraints remain. It has been often pointed out to me that I open myself to the criticisms made by earlier logicians such as Church and Carnap. According to those critiques, exchanges on the basis of the even stricter norms of (i) logical equivalence and (ii) necessary equivalence (inside a possible worlds semantics) lead to unintuitive consequences. How could then we allow here true reporting on the basis of seemingly looser equivalences?

Suppose that John believes that Walter Scott is human (I assume that Scott is human, is *actually* true to the facts and *necessarily* so). Then (i) by Church's trickster's use of the Kronecker function—thus by *logical* equivalence—John believes that Utah has 29 counties and (ii) by the possible worlds semantics identification of all necessities as one proposition, John also believes that water is hydrogen hydroxide (or any other necessity).

I believe both Church's and Carnap's worries are ill founded—the substitution failures on the basis of their (logical, necessary) *sentential* equivalences are not "paradoxes" but expected results. Neither logical nor necessary equivalence provide a *natural* basis for reporter's shifts (and the illusion that there should be such substitutions is an artifact of the artificial semantic representations—either model theoretic or proposition theoretic—that Carnap and Church had adopted). The Carnap-Church troubling exchanges are substantially different from ours. Here is why.

These classical cases operate at (i) the level of the *inter- sentential* equivalences and (ii) *abstract* intentionally from what the singular term or the predicate are *about* (what individual and property they pick out). Our examples have the opposite profile. They are focused on the particular constituents—objects, properties—picked out in the original thinker's speech. The reporter fixes, by other linguistic means, on *these very* objects and properties—he operates *de-re*. This is why the reporting is intuitively about the same topic—Godiva and her size features. This is also why, on the current semantics, arbitrary logical and necessary equivalence that *multiply through* these *intra*-sentential considerations, factor out the *res*, do not warrant reporting shifts.

We may well say: the classical "criteria" of inter-sentential logical and necessary equivalence are often (i) way too weak a condition and (ii) for many cases (like ours), way too strong (and irrelevant) a condition.

There is a long tradition—I will call it the *logical form tradition*—of having none of this form-transcendence of rationality. The logical form theorist argues as follows: let the two parameters of (i) data base and (ii) available logic, be indeed fixed and shared by Peter and Max, just as you said. Still, upon analysis and *true* form exposure, what we should see Peter as pondering and accepting is a claim of the noncontradictory *logical form* "*a* is *F* and *b* is not *F*"; what Max ponders and accepts is of the contradictory form "*a* is *F* and *a* is not *F*." Sure, the *visible* English sentences they ponder (accept) are form-indiscernible. But we can code into the genuine thoughts (logical forms) they contemplate the difference—what absolves Peter is that he *really* contemplates a noncontradictory form.

I beg to disagree about the revisionist Linguistics, Logic, and Cognitive theory (of rationality). The revisionist methodology provides yet another case where logical ingenuity subverts one away from honest toil (I would not call it "theft" but honest toil it is not). If "Paderewski is identical with Paderewski" is not of the "*a* is identical with *a*" form, I believe that no honest to God common currency English sentence is of the form "$a=a$" ("common currency" is meant to deflect us here from Russell's privately fabricated logically proper names for private internal goings on).

Now, I (and we all) do believe common currency English has such "$a=a$" sentences, for example, "London is London." And so, we have to refrain from serving ourselves to the revisionist linguistics/logic/rationality theory. The linguistics (syntax) of "Paderewski is Paderewski" is neither that of "Paderewski is identical with itself" nor that of "Paderewski is Hieronymous" (another name of that Pole). In turn, the logic of natural languages sentences is not as simple as that of predicate logic forms—"Paderewski is Paderewski" (true or false) is not an *instance* of the reflexive-involving "Everything is identical with itself." This impacts our notion of rationality. When our agents think in and are reported by English sentences, rationality may not be as *intrinsic* a feature as we would like it to be. Intuitively, Pierre and Peter are rational but Max is not. What makes us sure is not the sentences they assent to (or ours reporting them) but the *circumstances* in which they accepted/expressed the sentences.

Like remarks apply to indirect discourse reports. Sometimes when I report "*P* believes that *a* is *F* and that *a* is not *F*," I do point to the irrationality of *P*; sometimes not (for example, as with Peter/Pierre). Nothing in the *form* of the report settles the matter. The critical question is what the background story of the report is, as recorded "from above."

VI.3. The Content Dilemma and the Unification of Language and Thought. It is time to go back to our starting point, Frege's ideal of unifying our accounts of language and thought, his unifying (i) our *use* of the word 'London' to say (and to report that others said) that London is pretty and (ii) our *thinking* that it is pretty. Frege's project ran into serious trouble round 1970. A common diagnosis was that the troubling data indicated that the very ideal of unification should be given up. For the common diagnosis suggested that (i) the quartet's new data established that the semantics of our referential devices must be content-free but (ii) cognitive psychology and the analysis (and perhaps even attribution) of thoughts call essentially for reference-transcending contents. We thus ran into the *content dilemma.*

A common contemporary reaction to the dilemma had in effect accepted the pessimism inherent in its formulation. I grouped such unification-pessimistic reactions under the "more content" title, working as they do to generalize Frege's original content program, so that it has both (i) reference-bound contents sensitive to the needs of semantics, and (ii) reference-transcending contents-senses sensitive to the needs of psychological description. This method of "divide and conquer"—either in the form of the two-dimensional semantics of Stalnaker or in the character/content theory of Kaplan—conquers indeed a variety of puzzles in semantics and cognition; but it does divide apart semantics and cognitive psychology. Another aspect of this segregative reaction we confronted was the separation of *linguistic* questions—how we report cognition in ordinary language—from questions of true *cognitive structure*—what is the constitution of the thinkings proper. The segregation was encapsulated in the disunity thesis "reports are de re, beliefs (I would say: believings) are de dicto." One could not but conclude: so much the worse for the original unification project.

My own reaction to the 1970 data is quite different. Like Frege, I think we ought, can, and thus must pursue the unification project.

Like Frege, I think that, in carrying out the project, we should separate end and means. In this respect, the "more content" and disunity thesis reactions to the 1970 data seem to me to reverse cart and horse: we reject Frege's end—unification of semantics and cognition—but save his means—the content program. I suggested that the 1970 uprising teaches us the opposite lesson: we should pursue Frege's end but reject his means. Like Frege, I find the key to the prospects of the unification project not in this or that theoretical semantics, but in a better understanding of the primal cognitive facts—the Frege data in the first few lines of *SB*, the data in Frege's insightful *SB* footnote, and the Kripke (Donnellan, Putnam) "in the head data"—all simple *facts* about how we think of objects.

It is thus that I view the 1970 data—and indeed earlier anticipations by Frege—as primarily about cognition (with linguistic reference merely a special case): our cognition of objects is not as content-imbued as we so often presuppose. It is *because* our cognition is object-bound and content-free that our semantics is, in turn, purely referential.[32]

JOSEPH ALMOG

University of California at Los Angeles

---

[32] This paper has been slow in coming. One part—regarding the solution of Frege's informative identities puzzle—was completed in the late eighties and presented at various institutions (under the title "The First 13 Sentences of *Über Sinn und Bedeutung*"). It was much helped by conversations with Kaplan and especially Wettstein, who published at the time his independent dis-solution of the puzzle as "Cognitive Significance without Cognitive Content," *Mind*, XCVII (1988): 1–28, now further developed in his book, *The Magic Prism*. The extension to indirect discourse report-sentences, including what is called *Frege's footnote insight* (indirect discourse reports by denotation only), occurred to me only in the mid-nineties and was presented in various talks since then (as mentioned in footnotes above, my stand on indirect discourse reports has been much shaped by conversations with Eaker). The complex history of the paper has one further twist. While this paper was slowly taking shape, a twin—dealing with the other major puzzle reputed to threaten "direct reference," the problem of *empty names* (and no thought contents)—came into being, entitled "The Proper Form of Semantics" (first published in a Donnellan *festschrift* and reprinted in *Descriptions and Beyond*). It develops the idea (central also to the present paper) of giving semantic and cognitive descriptions from the vantage point of what Donnellan called "the omniscient observer of history."

STANDING ON COMMON GROUND

Contradictions, impossible objects, and worse may be hard to believe in and, thus, hard to accept. When we consider logic, the same apparent incredulity creeps in: How, to presage the discussion below, can we accept contradictions if it means jettisoning the *Law of Noncontradiction*? To be sure, even if one adopts a fairly permissive (Quinean) stance towards "deviant logics," one might, nonetheless, be baffled by particular acts of desertion—as no doubt Quine would. But being baffled by the desertion of a logical principle is different in kind from rejecting the possibility of the desertion of a logical principle on grounds of incoherence—by refusing to understand even how a debate about the principle could be set out.

In this paper, we introduce, and attempt to undermine, what we call *bedrock reasoning*. Bedrock reasoning—evident, most explicitly, in David Lewis's work,[1] but implicit in much philosophical theorizing—is both philosophically pervasive and, we maintain, philosophically unsupported. We aim to highlight bedrock reasoning, in order to undermine it as methodologically suspect.

I

In 1999, prior to the publication of *The Law of Non-Contradiction: New Philosophical Essays*, Lewis was invited to contribute a paper to that volume. In the course of declining, Lewis wrote:

> I'm sorry; I decline to contribute to your proposed book about the 'debate' over the law of non-contradiction. My feeling is that since this debate instantly reaches deadlock, there's really nothing much to say about it. To conduct a debate, one needs common ground; principles in dispute cannot of course fairly be used as common ground; and in this case, the principles *not* in dispute are so very much less certain than non-contradiction itself that it matters little whether or not a successful defense of non-contradiction could be based on them.[2]

---

[1] See, for example, "Logic for Equivocators," in his *Papers in Philosophical Logic* (New York: Cambridge, 1998), pp. 431–41; "Noneism or Allism?" in his *Papers in Metaphysics and Epistemology* (New York: Cambridge, 1999), pp. 152–63; and in his remarks in a letter quoted in the preface to Graham Priest, JC Beall, and Bradley Armour-Garb, *The Law of Non-Contradiction: New Philosophical Essays (New York: Oxford, 2005), p. 1.*

[2] *The Law of Non-Contradiction*, p. 1. The first sentence of this quotation—as literally stated—was falsified by its appearance in the preface to the volume. Part of the purpose of this paper is to indicate why the other sentences in the quotation are false, too.

0022-362X/05/0210/532–44

This—in its essentials, anyway—is a view Lewis had had for a long time. Back in 1982, he writes (in response to the view that some sentences are both true and false—true with true negations):

> No truth does have, and no truth could have, a true negation. Nothing is, and nothing could be, literally both true and false. This we know for certain, and a priori, and without any exception for especially perplexing subject matters. The radical case for [a logic presupposing this] should be dismissed just because the hypothesis it requires us to entertain is inconsistent.

He continues:

> That may be dogmatic. And it is: I am affirming the very thesis [called into question by my opponents] and—contrary to the rules of the debate—I decline to defend it. Further, I concede that it is indefensible against their challenge. They have called so much into question that I have no foothold on undisputed ground. So much the worse for the demand that philosophers always must be ready to defend their theses under the rules of debate.[3]

Elsewhen (in 1990), Lewis responds to Routley's suggestion that "there is neutral quantification," in addition to ontologically loaded quantification, where the former ranges over both existent and nonexistent entities alike, by saying that Routley

> purports to quantify without quantifying. For when he quantifies neutrally he is not quantifying in the one and only way there is to quantify, since ex hypothesi the one way is the loaded way. This we find altogether unintelligible.[4]

In all the above quotations, Lewis characterizes and defends his use of what we call *bedrock reasoning*:[5] the (*purportedly justified*) refusal to engage in a debate over the value of deserting a principle. When engaging in such reasoning, Lewis and others often invoke the bedrock quality of the principle they refuse to give up. But, as is usual with philosophers, the principle itself does almost no work to justify their intransigence towards changing it. In what follows we unearth the *rationales* behind Lewis's invocation of bedrock reasoning, and show why they are not compelling in the way he thinks.

---

[3] "Logic for Equivocators," p. 101.

[4] "Noneism or Allism?" p. 158.

[5] Wittgenstein, in his *Philosophical Investigations*, G.E.M. Anscombe, trans. (New York: Macmillan, 1958), p. 85^e, §217, writes: "If I have exhausted the justifications I have reached bedrock, and my spade is turned. Then I am inclined to say: 'This is simply what I do'."

Prior to doing so, we offer a few clarifying points. First, we distinguish two formulations of the Law of Noncontradiction—the *weak* version (hereon, the *LNC*) and the *strong* version (hereon, the *LNC**). By the weak version we mean: the principle according to which, for all meaningful sentences, *S*, $\sim(S \ \& \ \sim S)$. This principle—accepted by dialetheists and classical logicians alike—tells us, colloquially, that all contradictions are false. For the classical logician (who, in particular, rejects inconsistency)—the *consistentist*, as we will call her—this version of the LNC is uniformly accepted, its negation, assuming double-negation elimination—what we might call the *LC*—uniformly rejected.

The dialetheist accepts that some sentences are true with true negations. As the negation of some true sentences are also true, it follows that, for some sentence, *S*, and its negation, $\sim S$, the conjunction of these sentences, $(S \ \& \ \sim S)$, is, likewise, both true and false. As this conjunction is true, the relevant instance of the LNC is false; as this conjunction is false, the relevant instance of the Law of Noncontradiction is true. Thus, the LNC and the LC are both true and false—*true dialetheia*, two-way truths.

The consistentist and the dialetheist thus agree on the weak version of the LNC, although they disagree on the *status* of the relevant instances of the LC. But this is not the only point of disagreement. For, while the consistentist accepts the LNC, she takes it to follow from a stronger version of the LNC—the LNC*—which she, therefore, accepts.[6] The LNC* says that, for all meaningful sentences, *S*, $(S \ \& \ \sim S)$ is *false only*, where, being false only, she would thereby reject the claim that they are—even possibly—also sometimes true. While the dialetheist accepts the LNC, he does not accept all instances of the LNC*. Insofar as the LNC* is taken to generalize—to apply to *all* meaningful sentences—the dialetheist thereby rejects it, although, along with the consistentist, he endorses the LNC, as generalized. Thus, to be clear, when Lewis declines to get embroiled in a debate over the Law of Noncontradiction, it is the LNC*, not the LNC, that he refuses to discuss—for this principle, he maintains, cannot be up for debate.

II

It may seem that the weakness of Lewis's justification of his intransigence towards the possibility of debating the law of noncontradiction is right on the surface. After all, as we have seen, he (in his "Logic for Equivocators") invokes the a priori status of the law of noncontradiction; and aprioricity, in any form, is hardly philosophically uncontentious. Moreover, some such view of the a priori does seem linked to the position he takes on Routley's nonstandard quantification: where *I* stand, the logical idioms *I* employ, constrain what can be translated into *my* language, and in turn, constrain what *I* can understand. But this claim about understanding (or the failure of such), seems endangered by two straightforward facts: first, that—as Lewis well knows—there are cogent logical alternatives to the orthodox position on logical idioms; and second, that a rather large number of people seem to have no problems working with such alternatives, and even adopting them—understanding what such alternatives come to does not seem to be problematical.

---

[6] Of course, this is not the *only* reason for accepting the LNC, for the principle can be justified by logic, as well.

A contrast might help. The dialetheist, for example, is able to make quite clear what his dialetheism amounts to, and how he motivates it. Similarly, Routley is very clear on his Meinongian position. Indeed, what makes Lewis (in his "Noneism or Allism?") seem so disingenuous, in his refusal to understand Routley in the terms that Routley wishes to be understood, is that it does not seem hard to grasp what Routley is up to—in Routley's terms—at all! In order to mark the contrast between Lewis's refusal to understand and the admission of an ordinary failure to understand something, consider a family of theories that really do baffle the understanding. Recall, for example, what Richard Feynman says about QED:

> It is my task to convince you *not* to turn away because you don't understand it. You see, my physics students don't understand it either. That is because *I* don't understand it. Nobody does.[7]

Similar remarks can be, and indeed have been, made about quantum mechanics, for these are truly puzzling theories that *really do* appear to elude ordinary understanding. Dialetheism, however—whether you accept it or not—seems different from these cases. One may have trouble either believing or accepting dialetheism, but it by no means follows that the position is *hard to understand.* It may be that one can preserve a more pristine notion of "understanding"—by claiming that when one purports to understand alternative logical idioms, and the formal systems within which they flourish, one really is only mechanically manipulating such systems syntactically; and it can be denied that the mechanical manipulation of a nonorthodox logical apparatus to yield implications involves *genuine* understanding. It is possible, that is (if we connect this view with the one Lewis offers about a priori principles), to manipulate syntactically-consistent systems that desert a priori principles without, of course, understanding *what is supposedly said* with the resources available in such systems.

---

[7] In his *QED: The Strange Theory of Light and Matter* (Princeton: University Press, 1985), p. 9.

Let us briefly take up the concern just aired—that when one purports to understand alternative logical idioms, and the corresponding formal systems, one really is only manipulating such systems syntactically. It is worth noting in this regard that dialetheists, and proponents of other alternative logical systems, for that matter, take very seriously the provision of a Tarski-style semantics for such systems.[8] If the concern raised presupposes that a necessary condition for our understanding a formalism is that it be linked to a semantics in this sense, then dialetheism, in particular, meets the challenge.

One may worry, however, that such a semantics still will not help, as it presupposes the very alternative logical idioms that one fears that one does not understand. It is worth noting, at this point, that the debate has turned to the details of what understanding a logical idiom—any logical idiom—consists in. Worse, it appears that the debate has deadlocked, as a tool for providing an understanding of a formalism—semantics—is itself logically complex enough to be characterized in terms of—or without—the logical principles under contention.

To be sure, we have gone somewhat beyond Lewis's allusion to aprioricity, and suggest, in any case, that such allusions presuppose additional strong assumptions apart from the logical principles under contention. We want, therefore, to suggest that Lewis is on much firmer ground than his invocation of aprioricity and understanding suggests, and indeed, as the nonappearance of those notions in his (1999) letter, quoted above, may imply.[9] The way to do this is to consider a coherentist position—one somewhat like Quine's, say, where issues of "understanding" are not central[10]—and to show how the issue of the pointlessness of a debate over logical principles seems to emerge, on even the most radically construed coherentism.

---

[8] Priest, in his *In Contradiction*, part II, and chapter 9 of part III (Boston: Martinus Nijhoff, 1987), supplies such a semantics, and argues that dialetheism has advantages over the classical view because such semantics can occur in the very same language it is the semantics of.

[9] We are not deserting the issue of whether the purported understanding of alternative logical systems is only a matter of manipulating such systems syntactically. We return to this issue in section IV.

[10] Although we hasten to add: there are *many* coherentists, and *many* philosophers who have deserted a priori-conditioned notions of understanding; our invocation of *Quine* is merely iconic.

Imagine that one believes, along with Quine (at least in some of his moods[11]), that any statement whatsoever is open to revision. It is also part of Quine's position, however, and this is quite natural on other grounds, that logical principles—even though revisable—are peculiarly central. One way of understanding the import of this centrality—and a way that we will adopt for the purposes of this paper—is to take logical principles to be *topic neutral*: there is no subject matter to which they do not apply.[12] Now, even assuming this to be so, the coherentist faces a problem, if she wants to engage in a debate over whether a particularly central logical principle is one that should be revised. The problem is the one Lewis mentions—that so much is called into question that there seems to be no foothold on undisputed ground. Why no foothold? Because one cannot see how to *infer* implications—as one must—from one's claims in a neutral way so that two substantially different, and opposed, positions on *inference* can be compared.

Notice what the worry amounts to in this context. It is *not* one about the comprehensibility of alternatives, nor about principles being at issue that we cannot give up because of their (say) "foundational" or a priori nature. The problem is that one wants to compare alternative logics on *rational grounds*, where this is not meant in the silly contradictory sense that one wants to compare alternative logics on the basis of, among other things, the rational (that is, logical) principles *one already has*; but, rather, in the sense that one wants to, say, desert one logic for another for *reasons* which are convincing both to oneself and to others.

It is at this point that some will recall Carnap's (1950) views on linguistic frameworks: there are myriad alternatives, in languages we may adopt or logics that we may pursue. We are to try one or another and our decisions are made, when made, on pragmatic grounds of "efficiency, fruitfulness, and simplicity." These grounds, "although [themselves] not of a cognitive nature, will nevertheless usually be influenced by theoretical knowledge, just like any other deliberate decision concerning the acceptance of linguistic or other rules."[13]

To be sure, there is more than one interpretation possible of how we actually manage to make these pragmatic decisions. Carnap does suggest that such adoptions need not be (and in many cases, are not) deliberate. Still, one can claim, perhaps risking a vicious regress, that there is always *some* linguistic haven—a language (perhaps the ordinary one) where we can articulate the pragmatic considerations on which we base such decisions. Alternatively, and if we are wise enough to anticipate the threat of regress that this position faces, we can hope to justify (that is, to *reconstruct rationally*) our decision after the fact, in the language that we have subsequently chosen. That said, although there is no threat of vicious regress, if we do the latter, it seems that we are straightaway faced with the worry that such a rational reconstruction offers only *rubber-stamped* approval: we are, after all, going to use the very language *itself*, and *its* resources, to justify our choice of *it*. Such lovely options philosophy sometimes seems to offer: vicious regress or conflict of interest.

---

[11] For example, in his "Two Dogmas of Empiricism," in his *From a Logical Point of View* (Cambridge: Harvard, 1980, revised, 2nd ed.), pp. 20–46.

[12] An old (Kantian) term for this is that logical principles are "formal." See, for example, Gila Sher, "In Search of a Substantive Theory of Truth," this JOURNAL, CI, 1 (January 2004): 5–36, on pp. 30–31.

[13] All quotations are from p. 208 of Carnap's "Empiricism, Semantics, and Ontology," in his *Meaning and Necessity* (Chicago: University Press, 1950), pp. 205–21.

III

It is worth noting that debates over alternative logics, although they certainly invoke issues about ease and simplicity,[14] usually go well beyond such considerations. For example, the debate between the orthodoxy, and the particular heresy of dialetheism, turns on questions about what a language should or should not be able to say about its own properties (specifically, its semantic properties), about the value of self-referential (or circular styles of) reasoning, about which languages a language should or should not be able to translate the sentences of, and so on.[15] And an examination of the debate seems to reveal—at least on the surface—that the issues are being joined in an entirely nonquestion-begging way.

The point is this: even if Lewis is ultimately right that the debate will deadlock, with opponents on neither side yielding, this is not because the principles under dispute are so fundamental that the debate is merely in terms of whether such principles are to be accepted or rejected, so that questions are begged on both sides. Of course the beggings of questions need not be particularly *obvious*: more than one philosophical paper exists only because it revealed a begging of questions where no one presumed such a thing to be. Philosophical arguments occur in the murky domain of ordinary language, where presumptions are slickly buried stylistically—even despite the best of intentions of talented and subtle philosophers—and where, in any case, the inferential tools at hand are not definitively understood. Nevertheless, it does seem that if a debate between dialetheists and classical logicians joins over issues (say) about the expressibility capacities of ordinary English, and if the former argues both that natural languages are "semantically closed" and that this phenomenon cannot be captured except on a dialetheist approach (as Graham Priest, in *In Contradiction*, argues), then one may disagree, either because she thinks that the dialetheist is wrong about classical resources, or about the purported fact of semantic closure, or because she disagrees with the implications that Priest draws about these facts (for example, she thinks that English is inconsistent, not in the logically interesting way Priest suggests, but in the logically uninteresting way that Tarski[16] originally suggested). What is important to recognize, however, is that the argument does not devolve into anything question begging—at least not on the trivial ground that a tug of war over the LNC* has ensued.

---

[14] For example, recall Quine's panegyric about the "sweet simplicity" of classical logic in "What Price Bivalence?" in his *Theories and Things* (Cambridge: Harvard, 1981), p. 31–37, here p. 32. Of course, he does more to argue for classical logic's simplicity in his *Philosophy of Logic* (Englewood Cliffs, NJ: Prentice-Hall, 1970).

[15] See, for example, Armour-Garb, "Wrestling with (or without) Dialetheism," *Australasian Journal of Philosophy*, LXXXIII (2005): 87–102; Azzouni, "The Strengthened Liar, the Expressive Strength of Natural Languages, and Regimentation," *The Philosophical Forum*, XXXIV, 3–4 (2003): 329–50; Beall and Armour-Garb, "Should Deflationists be Dialetheists?" *Noûs*, XXXVII (2003): 303–24; Priest, *In Contradiction*; Priest, "What Is So Bad about Contradictions?" this JOURNAL, XCV, 8 (August 1998): 410–26.

Here, however, we seem to run up against Lewis's claim, quoted earlier, that "the principles not in dispute are so very much less certain than non-contradiction itself...." But that claim seems especially unfair to raise at this juncture. For here is a debate, looking quite like other familiar debates, as they ordinarily arise in philosophy: assumptions are made, deployed against opponents, teased out and refined, and so on. Lewis then comes along, and offers a genuinely question-begging argument: Why engage in a debate over the LNC*, since, after all, we are more sure of it than of any of what has been raised as considerations in this debate?

The foregoing description of the nature of this sort of debate seems to suggest that Lewis is wrong at least in the following sense: debates over logical principles—in particular, though not necessarily limited to the LNC*—need not, on the surface, be question begging, as evaluating alternative approaches need not officially turn on the invocation of the principles that are at issue.

It may seem, however, that this can only be a matter of sheer appearance. After all, if the principle of the LNC*, to take an example, is central to our reasoning in the way that, *ex hypothesi*, nearly everyone takes it to be, then how can this debate occur *without* that principle being deployed, surreptitiously, by one side or—since dialetheists are either explicitly or implicitly using the same ordinary inferential tools that everyone else is using too—by both sides?[17]

---

[16] "The Concept of Truth in Formalized Languages," in John Corcoran, ed., *Logic, Semantics, Metamathematics* (Indianapolis: Hackett, 1983), pp. 152–278.

A way out of this conundrum is not far to seek, if we recall the earlier characterization of logic as "topic neutral." There are many ways to characterize logical principles[18] but, ultimately, they must turn out to govern (or be composed of) sentences couched in the vocabulary of the language in which they occur. They may transcend that language and its vocabulary, in one or another way: it may be that they are understood to immediately apply to new vocabulary, when coined in a language; they may also, strictly speaking, be items that literally quantify over the sentences of the language (and in this way indicate the instances of the language that exemplify them); or they may be schemata. How, exactly, this goes does not matter, for, in a sense that cannot be avoided, they are generalizations, and, in describing them as topic neutral, one is simply claiming that there are no restrictions upon them—that is, instances of them may involve terminology of any sort.

In just these terms, then, a desertion of a logical principle can be characterized: It is a desertion of the topic neutrality of a logical principle, its demotion of range, from terminology of the entire language to some restricted subset thereof.[19] If this is right, then it explains away our conundrum—explains how it is possible to argue about a logical principle. That is, it explains how to argue against one another in a nonquestion-begging way about that logical principle, and how to do so, even while using *instances of* that logical principle.

---

[17] Here is a simple-looking proof that both proponents of classical logic and dialetheists are using the same inferential principles: *pace* Lewis, *we all understand each other's arguments.* Anyway, we understand each other's arguments as well (or ill) as we understand those of any other philosopher—proponent of classical logic, or otherwise. Priest, in his *In Contradiction*, for example, does not seem to be running particularly *obscure* arguments *at all.*

[18] One way to do so may be found in Azzouni, "A Priori Truth," *Erkenntnis*, XXXVII (1992): 329–50, here pp. 329–30.

[19] Three points. First, "some subset thereof" may be understood as the empty set (so that all the instances of the principle are deserted). This however, is not the standard case: it is not the case with respect to the logical alternatives—quantum logic, dialetheism, intuitionism—that are the most robust alternatives to classical logic. Second, this is precisely the way that the dialetheist's desertion of the LNC* should be, and is, understood: the principle has been demoted from a topic-neutral principle that we may simply take for granted in any context, to one that holds only in specific circumstances, so that we may actually employ the principle, provided we know that such circumstances hold. Third, this methodology of focusing on the *instances* of a logical principle held constant—or some privileged set of principles, for that matter—instead of the topic-neutral generalization of it, to explain how rational adjudication of conceptual change is possible (and on what grounds), is one that Azzouni has long championed. See his *Metaphysical Myths, Mathematical Practice: The Ontology and Epistemology of the Exact Sciences* (New York: Cambridge, 1994), part 3, §4, or his *Knowledge and Reference in Empirical Science* (New York: Routledge, 2000), part II, §2.

For, first, to use instances of a logical principle need not be to accept *all* of its instances. Second, in arguing against a logical principle, one is, generally speaking, arguing not for its wholesale elimination—an elimination of *all* its instances, or, if you will, all its "applications"—but only for a subset of such, which does not include (or need not include) the applications of it to the very context of the argument itself. Third, the instances of a logical principle that both parties agree to are precisely (part of) the common ground in terms of which both parties can join the debate. And, fourth, after the fact, each party may be still able to describe the reasoning of *both* contestants *in their own terms*, and in a way that nevertheless still begs no questions against the opponents.

So, for example, imagine that *B* argues for the demotion of LNC* from topic neutrality to the status of a principle that is to apply only when certain sorts of self-reference do not occur. As long as those certain sorts of self-reference do not occur in the reasoning that *B* uses to make his or her case, *B* can employ instances of the LNC*.

In turn, *J*, a supposed proponent of the LNC*, may describe *B*'s arguments using exactly the instances of the LNC* that *B* invoked (supposing that *B* really did so invoke such instances) and by noting that *B* justifies these uses by citing the special circumstances that allow those instances to be used in this case, rather than on sheer logical grounds. *B*, on the other hand, can describe *J*'s argument, noting that *J* uses instances of the LNC*. But *B* can note, in turn, that from *B*'s perspective, these uses are fine, since special circumstances—which circumstances *J* did not worry about, since *J* invoked it as a logical principle—obtain. The reasons—that is, logical and nonlogical principles—that *B* and *J use*, in characterizing the arguments of their respective opponents (in addition to their own), are not the same, to be sure, but the *instances* of those reasons are the same. More accurately: the principles that they invoke in each case may actually be the same but, in that case, the *status* of the principles is not the same.[20]

It is worth, at this point, lodging a specific objection to Lewis's take on the debate about the status of the LNC*. It is peculiar, actually, to treat a logical principle, when it is topic neutral, as something a priori or obvious; and this is because such a principle has such sweeping scope. Once bizarre epistemological views—about how one could acquire a grasp of such principles, for example—are excluded, one is forced to a position something like this: the obviousness of a logical principle is established on the basis of very natural and common applications of it.[21] The generalization of these applications, to all instances of the principle, is easy to make, but risky—especially because it goes so far beyond any applications that we have experienced or that we can even imagine. It is better to recognize that the obviousness of logic resides only in the specific *instances* of it which we find obvious or indispensable, and not in the topic-neutral construal of logical principles, which *intuition* illegitimately extends it to. Therefore, the common ground of argument is precisely those instances that we find so obvious, and not the principle itself.

---

[20] We are not arguing that debates over logical principles are *never* question begging. We are arguing that such debates *need not* be question begging, and we are illustrating how such a case can occur.

The analogy with empirical generalizations is tight: an (Aristotlean) law, which tells us that objects set in rectilinear motion eventually stop, when seen as a law, is seen as holding of *all* rectilinear motions. Its falsification does not amount to a denial that objects set in rectilinear motion *ever* stop; it is simply that the generalization mutates into one that applies only under special circumstances. For this reason, evidential arguments used to refute that law—as a law—may still help themselves to data, from circumstances in which the special conditions hold (and so, where objects set in rectilinear motion soon stop). The situation is, we maintain, no different in logic.

IV

Let us return briefly to the issue of understanding alternative logical idioms, as promised in section II.[22] An issue about the nature of our *understanding* such alternatives is a more delicate matter than the one on which we have been focused, as it turns on the details of the alternatives in question. Nevertheless, we have already noted that the worry that alternative logical systems must be manipulated syntactically without understanding is undermined—at least to the extent that semantics for such alternative systems are generally available; and, on many views, the availability of such semantics encapsulates what "understanding" comes to.

---

[21] See *Metaphysical Myths* (part 3, §4), where just this explanation for a priori intuitions about logic and mathematics is offered.

[22] We note that, strictly speaking, the present issue is not central to the topic of the paper, which has been the question of whether debates over fundamental logical principles can be joined in a nonquestion-begging way. Our thanks to the editors of this JOURNAL for, nevertheless, inviting us to say something more about it.

But we think we can press the argument further in light of the focus on instances of a generalization that has been explored in the foregoing sections, and we will do so with respect to the special case of the desertion of the LNC*. The claim is this: our sense that we understand a logical principle battens not on generalizations—not even on semantic generalizations. We believe this if only because—as mentioned in section II—such semantic generalizations rely on the same logical principles of which, on this view, they are supposed to underwrite the understanding. Instead, and noncircularly, our sense that we understand a logical principle turns on the instances of it that we are comfortable with.

This does suggest, of course, that if we were to try to adjudicate between two logical alternatives that had no logical principles with at least some instances in common, there would indeed be no common ground, and no common ground of *understanding*. But that is not the nature of the debates about logical principles that we are familiar with—in particular, it does not apply to the debate over the LNC*.

The worry that arises, therefore, is not about the instances of the LNC* that the dialetheist and the consistentist *agree on*, and it is not, as we cast it in section II, that the dialetheist position itself is not comprehensible. Rather, the worry is that one finds incomprehensible the very instances of (S & $\sim S$) that one is to regard as both true and false (on the dialetheist position).

We want to suggest that how incomprehensible one finds such instances turns rather directly on exactly which instances are under discussion. It is not irrelevant—not a coincidence—that the instances for which the dialetheist wants to desert the LNC* are among ones that some already regard us as not quite "understanding." Consider the Liar sentence. We suggest that one can see very well *how*, given what a Liar sentence *means*, it can be both true and false—and that it is very simple reasoning that can be used to establish how this is so. That simple reasoning tracks our understanding of how a Liar sentence can be both true and false—it does so, anyway, as well as we may be taken to understand such a sentence to begin with.[23]

But one—again, speaking intuitively—is prone to step back and say: *But how can this be?* We have to understand this question narrowly—

---

[23] And how well is that? Not badly, we would claim: we understand what the terms appearing in such a sentence refer to—after all, such understanding drives the paradox; and we understand how to draw the conclusion that the sentence is true if and only if it is false. Notice furthermore, that our understanding of how the Liar paradox could be both true and false is driven by an appreciation of the *semantics* of the sentence—including, of course, what the terms in it refer to.

not as one that presupposes the LNC*. (We are not understanding "*But how can this be?*" as blatantly begging the question.) Rather, such a question must presuppose something else, for example, that every indicative sentence of such-and-such a sort must correspond to a fact about the world, or something like that. If one is thinking this way, it may be that one will find the suggestion that a sentence can be both true and false to be incomprehensible.[24] But with respect to self-referential sentences such as the Liar, one can take a very different line. Truth and falsity of (particular) sentences are not always matters of correspondence—and self-referential sentences such as Liars are primary illustrations of this.

CONCLUDING REMARKS

Rationality, such as it is, is not a matter of general topic-neutral principles that one is driven to invoke whenever one offers an argument. Were that the case, we would be locked into our logical principles forever, unless we were capable of periodic Kuhnean leaps of faith to alternative principles—and were comfortable with regularly changing our views without being able to justify such changes. No, instead, rationality lies in the applications—the *instances*—of the logical principles we use. These applications allow us not only to hedge and refine those principles, outside the specific applications that are actually indispensable to us (if there is a good reason to do so), but to be able to provide *reasons* for hedging and refining such logical principles, on the basis of their instances, which will be held constant across the change contemplated. Exactly this is what allows a debate between dialetheists and the proponents of classical logic to be fruitful, rather than question begging (in the way that Lewis so glumly predicted); and it is what allows such opponents (on occasion) to even write papers *together* on the topic.

JODY AZZOUNI

Tufts University

BRADLEY ARMOUR-GARB

State University of New York at Albany

---

[24] And some dialetheists are willing to describe parts of the world—in some sense—as so contradictory, as corresponding, that is, to both the true and the false (as it were). But our point is that such a difficult-to-understand doctrine need *not* accompany a commitment to dialethism.

NEW BOOKS

KIM, JAEGWON: *Physicalism, or Something Near Enough.* Princeton: University Press, 2005. xiii, 186 p. Cloth $26.95.

KNUUTTILA, SIMO: *Emotions in Ancient and Medival Philosophy.* New York: Oxford, 2004. x, 341 p. Cloth $74.00.

KOCHAR-LINDGREN, GRAY: *Technologics: Ghosts, the Incalculable, and the Suspension of Animation.* Albany: SUNY, 2005. ix, 222 p. Cloth $75.50, paper $24.95.

KUSPIT, DONALD: *The End of Art.* New York: Cambridge, 2004. xi, 208 p. Cloth $28.00, paper $18.99.

LANG, BEREL: *Post-Holocaust: Interpretation, Misinterpretation, and the Claims of History.* Bloomington: Indiana UP, 2005. xviii, 200 p. Paper $45.00.

LAWLESS, ANDREW: *Plato's Sun: An Introduction to Philosophy.* Toronto: University Press, 2005. xiii, 364p. Cloth $65.00, paper $29.95.

LEAR, JONATHAN: *Freud.* New York: Routledge, 2005. xvi, 278 p. Paper $22.95.

LECHTE, JOHN, and MARIA MARGARONI: *Julia Kristeva: Live Theory.* New York: Continuum, 2005. vi, 182 p. Cloth $89.95, paper $19.95.

LEE, DANIEL E.: *Freedom versus Intervention: Six Tough Cases.* Lanham, MD: Rowman and Littlefield, 2005. xix, 133 p. Cloth $60.00, paper $21.95.

LEE, MI-KOUNG: *Epistemology after Protagoras: Responses to Relativism in Plato, Aristotle, and Democritus.* New York: Oxford, 2005. x, 291 p. Cloth $74.00.

LEPORE, ERNEST, and KIRK LUDWIG: *Donald Davidson: Meaning, Truth, Language, and Reality.* New York: Oxford, 2005. xviii, 446 p. Cloth $65.00.

LEVI, ISAAC: *Mild Contraction.* New York: Oxford, 2004. xi, 245 p. Cloth $55.00.

LEWIS, MICHAEL: *Heidegger and the Place of Ethics.* New York: Continuum, 2005. xiv, 212 p. Cloth $105.00.

LEWIS, THOMAS A.: *Freedom and Tradition in Hegel: Reconsidering Anthropology, Ethics, and Religion.* Notre Dame: University Press, 2005. xiii, 258 p. Cloth $32.50.

LINGIS, ALPHONSO: *Body Transformation: Evolutions and Atavisms in Culture.* New York: Routledge, 2005. 154 p. Cloth $95.00, paper $22.95.

LIVINGSTON, PAISLEY: *Art and Intention: A Philosophical Study.* New York: Oxford, 2005. xviii, 251 p. Cloth $55.00.

LOSEE, JOHN: *Theories on the Scrap Heap.* Pittsburgh: University Press, 2005. viii, 206 p. Paper $24.95.

LÜTZEN, JESPER: *Mechanistic Images in Geometric Form: Heinrich Hertz's Principles of Mechanics.* New York: Oxford, 2005. xiii, 318 p. Cloth $154.50.

MACDONALD, CYNTHIA: *Varieties of Things: Foundations of Contemporary Metaphysics.* Malden, MA: Blackwell, 2005. x, 278 p. Paper $29.95.

MACEY, DAVID: *Michel Foucault.* London: Reaktion, 2004. 160 p. Paper $16.95.

MACINTOSH, KERRY LYNN: *Illegal Beings: Human Clones and the Law.* New York: Cambridge, 2005. xiii, 272 p. Cloth $28.00.

MANTZAVINOS, C.: *Naturalistic Hermeneutics.* New York: Cambridge, 2005. xv, 180 p. Cloth $65.00.

MARKS, JONATHAN: *Perfection and Disharmony in the Thought of Jean-Jacques Rousseau.* New York: Cambridge, 2005. viii, 191 p. Cloth $60.00.

MATTHEN, MOHAN: *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception.* New York: Oxford, 2005. xxii, 362 p. Cloth $74.00.

MAY, LARRY: *Crimes against Humanity: A Normative Account.* New York: Cambridge, 2005. xiii, 310 p. Cloth $70.00, paper $25.99.

MAY, TODD: *Gilles Deleuze: An Introduction.* New York: Cambridge, 2005. ix, 184 p. Cloth $60.00, paper $19.99.

MCCARTHY, JOHN, and PETER WRIGHT: *Technology as Experience.* Cambridge: MIT, 2004. xi, 211 p. Cloth $35.00.

MCCUMBER, JOHN: *Reshaping Reason: Toward a New Philosophy.* Bloomington: Indiana UP, 2005. xix, 263 p. Cloth $24.95.

MCKEON, RICHARD: *Selected Writings of Richard McKeon, Volume 2: Culture, Education, and the Arts.* ZAHAVA K. MCKEON and WILLIAM G. SWENSON, eds. Chicago: University Press, 2005. xxiii, 362 p. Cloth $50.00.

MCKOWN, DELOS B.: *Behold the Antichrist: Bentham on Religion.* Amherst, NY: Prometheus, 2005. 417 p. Cloth $32.00.

MEHL, PETER J.: *Thinking Through Kierkegaard: Existential Identity in a Pluralistic World.* Urbana: Illinois UP, 2005. xii, 177 p. Cloth $35.00.

MENDELSOHN, RICHARD L.: *The Philosophy of Gottlob Frege.* New York: Cambridge, 2005. xix, 226 p. Cloth $75.00.

MILLAR, ALAN: *Understanding People: Normativity and Rationalizing Explanation.* New York: Oxford, 2004. xvi, 262 p. Cloth $74.00.

MILLER, MARK: *Philosophical Chaucer: Love, Sex, and Agency in the Canterbury Tales.* New York: Cambridge, 2004. x, 289 p. Cloth $75.00.

MILLGRAM, ELIJAH: *Ethics Done Right: Practical Reasoning as a Foundation for Moral Theory.* New York: Cambridge, 2005. vii, 360 p. Cloth $70.00, paper $27.99.

MONTIGLIO, SILVIA: *Wandering in Ancient Greek Culture.* Chicago: University Press, 2005. 288 p. Cloth $50.00.

MORGAN, VANCE G.: *Weaving the World: Simone Weil on Science, Mathematics, and Love.* Notre Dame: University Press, 2005. xi, 234 p. Cloth $55.00, paper $25.00.

MULHALL, STEPHEN: *Philosophical Myths of the Fall.* Princeton: University Press, 2005. viii, 126 p. Cloth $29.95.

MULLIN, AMY: *Reconceiving Pregnancy and Childcare.* New York: Cambridge, 2005. ix, 214 p. Cloth $70.00, paper $24.99.

MURPHEY, MURRAY G.: *C.I. Lewis: The Last Great Pragmatist.* Albany: SUNY, 2005. x, 477 p. Cloth $35.00.

MURPHY, PETER F.: *Feminism and Masculities.* New York: Oxford, 2004. xv, 288 p. Paper $29.95.

NEWTON, LISA H.: *Business Ethics and the Natural Environment.* Malden, MA: Blackwell, 2005. xiv, 259 p. Cloth $70.00, paper $29.95.

NICHOLS, SHAUN: *Sentimental Rules: On the Natural Foundations of Moral Judgement.* New York: Oxford, 2004. xi, 226 p. Cloth $55.00.

United States Postal Service

Statement of Ownership, Management, and Circulation

1 Publication Title: THE JOURNAL OF PHILOSOPHY

2. Publication Number: 0 0 2 2 - 3 6 2 X

3. Filing Date: 9/27/2005

4 Issue Frequency: MONTHLY

5. Number of Issues Published Annually: 12

6 Annual Subscription Price: Individual: $ 45 Institution: $100

7 Complete Mailing Address of Known Office of Publication (Not printer) (Street, city, county, state, and ZIP+4)
The Journal of Philosophy, Columbia University
1150 Amsterdam Avenue, MC 4972
New York, NY 10027

Contact Person: Pam Ward

Telephone: (212) 666-4419

8. Complete Mailing Address of Headquarters or General Business Office of Publisher (Not printer)
SAME

9 Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor (Do not leave blank)

Publisher (Name and complete mailing address)
The Journal of Philosophy, Inc.
1150 Amsterdam Avenue, MC 4972
New York, NY 10027

Editor (Name and complete mailing address)
Arthur Danto, President
The Journal of Philosophy, Inc., 1150 Amsterdam Avenue, MC 4972
New York, NY 10027

Managing Editor (Name and complete mailing address)
John Saylie, Managing Editor
The Journal of Philosophy, Inc., 1150 Amsterdam Avenue, MC 4972
New York, NY 10027

10. Owner (Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)

Full Name	Complete Mailing Address
The Journal of Philosophy, Inc.	Columbia University
	1150 Amsterdam Avenue
	MC 4972
	New York, NY 10027

11 Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → ☒ None

Full Name	Complete Mailing Address

12 Tax Status (For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
☒ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months (Publisher must submit explanation of change with this statement)

13. Publication Title: The Journal of Philosophy

14. Issue Date for Circulation Data Below: September 2005

15. Extent and Nature of Circulation			Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies (Net press run)			3500	3500
b. Paid and/or Requested Circulation	(1)	Paid/Requested Outside-County Mail Subscriptions Stated on Form 3541. (Include advertiser's proof and exchange copies)	700	666
	(2)	Paid In-County Subscriptions Stated on Form 3541 (Include advertiser's proof and exchange copies)	2400	2220 -
	(3)	Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Non-USPS Paid Distribution		
	(4)	Other Classes Mailed Through the USPS		
c. Total Paid and/or Requested Circulation [Sum of 15b. (1), (2),(3),and (4)] ▶			3100	2886
d. Free Distribution by Mail (Samples, compliment ary, and other free)	(1)	Outside-County as Stated on Form 3541		
	(2)	In-County as Stated on Form 3541		
	(3)	Other Classes Mailed Through the USPS		
e. Free Distribution Outside the Mail (Carriers or other means)			150	150
f. Total Free Distribution (Sum of 15d. and 15e.) ▶			150	150
g. Total Distribution (Sum of 15c. and 15f) ▶			3250	3036
h. Copies not Distributed			250	464
i. Total (Sum of 15g. and h.) ▶			3500	3500
j. Percent Paid and/or Requested Circulation (15c. divided by 15g. times 100)			95.38	95.06

16. Publication of Statement of Ownership
☒ Publication required. Will be printed in the 9/2005 issue of this publication. ☐ Publication not required

17 Signature and Title of Editor, Publisher, Business Manager, or Owner: John Saylie, Managing Editor

Date: 9/27/05

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

Instructions to Publishers

1 Complete and file one copy of this form with your postmaster annually on or before October 1. Keep a copy of the completed form for your records.

2. In cases where the stockholder or security holder is a trustee, include in items 10 and 11 the name of the person or corporation for whom the trustee is acting. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing corporation. In item 11, if none, check the box. Use blank sheets if more space is required.

3 Be sure to furnish all circulation information called for in item 15. Free circulation must be shown in items 15d, e, and f.

4 Item 15h., Copies not Distributed, must include (1) newsstand copies originally stated on Form 3541, and returned to the publisher, (2) estimated returns from news agents, and (3), copies for office use, leftovers, spoiled, and all other copies not distributed.

5 If the publication had Periodicals authorization as a general or requester publication, this Statement of Ownership, Management, and Circulation must be published; it must be printed in any issue in October or, if the publication is not published during October, the first issue printed after October.

6 In item 16, indicate the date of the issue in which this Statement of Ownership will be published.

7 Item 17 must be signed.

Failure to file or publish a statement of ownership may lead to suspension of Periodicals authorization.

You can now subscribe to the

THE JOURNAL OF
PHILOSOPHY

using

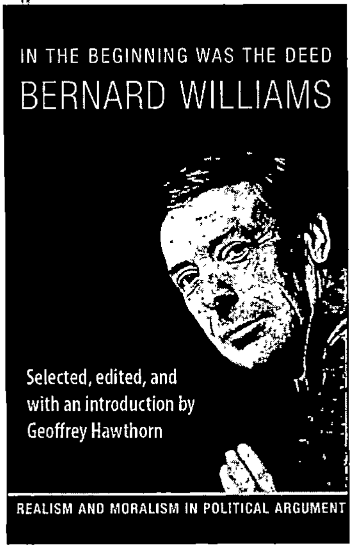
PayPal

New Rates January 2006

Institutional	$100
Individual	$45
Student/Retired	$20

To learn more about this easy new
option, click on the PayPal icon at
www.jphil.org

Insightful


IN THE BEGINNING WAS THE DEED
BERNARD WILLIAMS
Selected, edited, and
with an introduction by
Geoffrey Hawthorn
REALISM AND MORALISM IN POLITICAL ARGUMENT

IN THE BEGINNING WAS THE DEED

Realism and Moralism in Political Argument

Bernard Williams

Selected, edited, and introduced by Geoffrey Hawthorn

Bernard Williams is remembered as one of the most brilliant and original philosophers of the past fifty years. Widely respected as a moral philosopher, Williams began to write about politics in a sustained way in the early 1980s. This new collection of essays, most of them previously unpublished, addresses central subjects of political philosophy. A main theme throughout is that political philosophers need to engage more directly with the realities of political life, not simply with the theories of other philosophers.

"Bernard Williams saw the moral dimensions of politics more deeply and more clearly than any thinker of his generation."
—Ian Shapiro, Yale University

Cloth $29.95 ISBN 0-691-12430-2

New in paperback

WITTGENSTEIN ON THE ARBITRARINESS OF GRAMMAR

Michael N. Forster

What is the nature of a conceptual scheme? Are there alternative conceptual schemes? If so, are some more justifiable or correct than others? The later Wittgenstein already addresses these fundamental philosophical questions under the general rubric of "grammar" and the question of its "arbitrariness"—and he does so with great subtlety. This book explores Wittgenstein's views on these questions.

Paper $16.95 ISBN 0-691-12391-8 Due November

100 *Celebrating 100 Years of Excellence*
PRINCETON
University Press
800-777-4726
Read excerpts online
www.pup.princeton.edu

CAMBRIDGE

OUTSTANDING SCHOLARSHIP

Discovering Cell Mechanisms
The Creation of Modern Cell Biology
William Bechtel
Cambridge Studies in Philosophy and Biology
$80.00: Hardback: 0-521-81247-X: c.350pp

The Grammar of Society
The Nature and Dynamics of Social Norms
Cristina Bicchieri
$70.00*: Hardback: 0-521-57372-6: c.272pp
$24.99*: Paperback: 0-521-57490-0

Fichte: The System of Ethics
Edited by Daniel Breazeale and Günter Zöller
Cambridge Texts in the History of Philosophy
$75.00: Hardback: 0-521-57140-5: 447pp
$29.99: Paperback: 0-521-57767-5

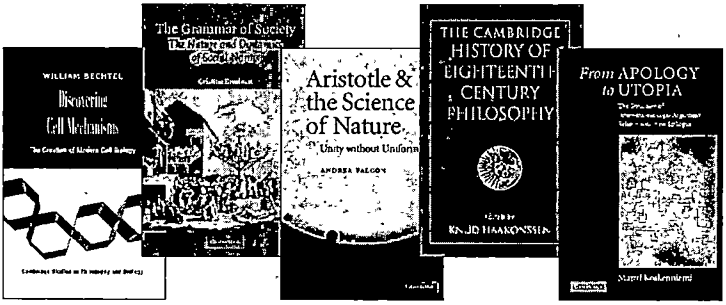
Friedrich Schleiermacher: Between Enlightenment and Romanticism
Richard Crouter
$80.00: Hardback: 0-521-80590-2: c.233pp

God and the Ethics of Belief
New Essays in Philosophy of Religion
Edited by Andrew Dole and Andrew Chignell
$60.00: Hardback: 0-521-85093-2: 294pp

Aristotle and the Science of Nature
Unity without Uniformity
Andrea Falcon
$75.00: Hardback: 0-521-85439-3: 158pp

The Cambridge History of Eighteenth-Century Philosophy
Edited by Knud Haakonssen
$140.00*: Hardback: 0-521-41854-2: c.968pp

From Apology to Utopia
The Structure of International Legal Argument
Martti Koskenniemi
$90.00*: Hardback: 0-521-83806-1: c.648pp
$36.99*: Paperback: 0-521-54696-6


Discovering Cell Mechanisms
The Grammar of Society
Aristotle & the Science of Nature
The Cambridge History of Eighteenth-Century Philosophy
From Apology to Utopia

*Prices subject to change.

www.cambridge.org/us/philosophy/

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

FROM CAMBRIDGE

Slavery, Philosophy, and American Literature, 1830-1860
Maurice S. Lee
Cambridge Studies in American Literature and Culture 148
$75.00: Hardback: 0-521-84653-6: 232pp

Illegal Beings
Human Clones and the Law
Kerry Lynn Macintosh
$28.00: Hardback: 0-521-85328-1: 286pp

The Cambridge Companion to Karl Rahner
Edited by Declan Marmion and Mary E. Hines
Cambridge Companions to Religion
$70.00: Hardback: 0-521-83288-8: 334pp
$25.99: Paperback: 0-521-54045-3

Epicurus on Freedom
Tim O'Keefe
$70.00: Hardback: 0-521-84696-X: 186pp

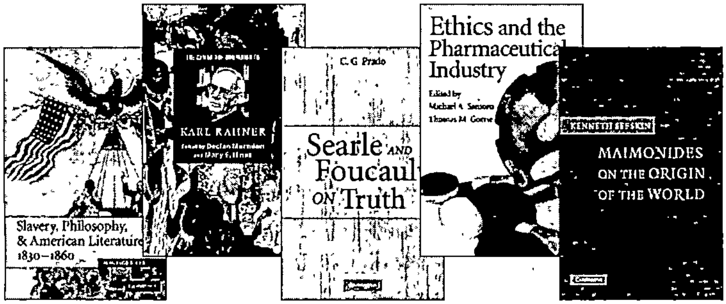
Searle and Foucault on Truth
C.G. Prado
$65.00: Hardback: 0-521-85523-3: 200pp
$23.99: Paperback: 0-521-67133-7

Ethics and the Pharmaceutical Industry
Edited by Michael A. Santoro and Thomas M. Gorrie
$40.00: Hardback: 0-521-85496-2: 528pp

Maimonides on the Origin of the World
Kenneth Seeskin
$55.00: Hardback: 0-521-84553-X: 224pp

Continental Philosophy of Social Science
Hermeneutics, Geneology and Critical Theory from Ancient Greece to the Twenty-First Century
Yvonne Sherratt
$70.00: Hardback: 0-521-85469-5: 256pp
$24.99: Paperback: 0-521-67098-5

Mary Astell
Theorist of Freedom from Domination
Patricia Springborg
$80.00: Hardback: 0-521-84104-6: 392pp

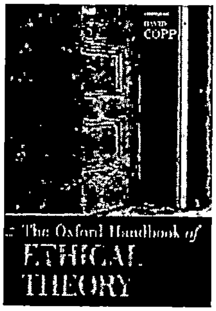

Slavery, Philosophy, & American Literature 1830–1860
KARL RAHNER
C.G Prado
Searle and Foucault on Truth
Ethics and the Pharmaceutical Industry
KENNETH SEESKIN
MAIMONIDES ON THE ORIGIN OF THE WORLD

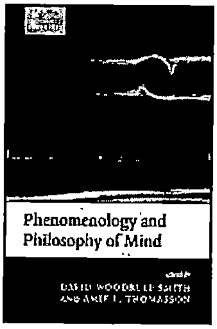
**Prices subject to change*

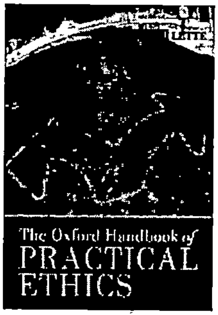
www.cambridge.org/us/philosophy/

CAMBRIDGE
UNIVERSITY PRESS

New from
Oxford




Phenomenology and Philosophy of Mind
DAVID WOODRUFF SMITH AND AMIE L. THOMASSON



Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.us.oup.com/us

OXFORD
UNIVERSITY PRESS

The Oxford Handbook of Ethical Theory

EDITED BY DAVID COPP

The Oxford Handbook of Ethical Theory is a major new reference work in ethical theory consisting of commissioned essays by leading moral philosophers.
(Oxford Handbooks)

November 2005 $74.00

Artha

Meaning

JONARDON GANERI

This second volume in the Foundations of Philosophy in India series is an important contribution to the philosophy of language. Here Jonardon Ganeri highlights the significant relationship between semantic power and epistemic power to understand the important philosophical category of meaning.

November 2005 $29 95

Reflections on Meaning

PAUL HORWICH

Paul Horwich's main aim in *Reflections on Meaning* is to explain how mere noises, marks, gestures, and mental symbols are able to capture the world—that is, how words and sentences (in whatever medium) come to *mean* what they do, to *stand* for certain things, to be true or false of reality. His answer is a groundbreaking development of Wittgenstein's idea that the meaning of a term is nothing more than its *use*.

January 2006 paper $26 95 cloth $74 00

New in Paperback!

The Oxford Handbook of Practical Ethics

Edited by HUGH LAFOLLETTE

The Oxford Handbook of Practical Ethics is a lively and authoritative guide to current thought about ethical issues in all areas of human activity—personal, medical, sexual, political, legal, and beyond.
(Oxford Handbooks)

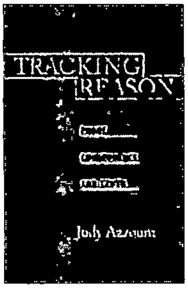
2003 (paper November 2005) paper $45.00 cloth $195.00

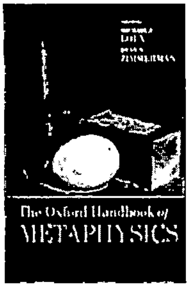
Phenomenology and Philosophy of Mind

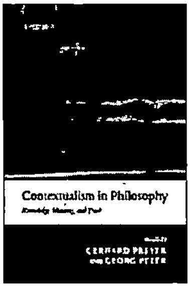
EDITED BY DAVID WOODRUFF SMITH AND AMIE L. THOMASSON

Philosophical work on the mind flowed in two streams through the 20th century: phenomenology and analytic philosophy. This volume aims to bring them together again, by demonstrating how work in phenomenology may lead to significant progress on problems central to current analytic research, and how analytical philosophy of mind may shed light on phenomenological concerns.

November 2005 paper $35 00 cloth $95 00







New from
Oxford

Tracking Reason
Proof, Consequence, and Truth
JODY AZZOUNI

Jody Azzouni investigates the connection between the ordinary notion of consequence and the formal analogues invented by logicians. One claim of the book is that, despite our apparent intuitive grasp of consequence, we do not introspect rules by which we reason, nor do we grasp the scope and range of the domain, as it were, of our reasoning.

November 2005 $49.95

Understanding Eating Disorders
Conceptual and Ethical Issues in the Treatment of Anorexia and Bulimia Nervosa
SIMONA GIORDANO

Simona Giordano presents the first full philosophical study of ethical issues in the treatment of anorexia and bulimia nervosa. Beginning with a comprehensive analysis of these conditions and an exploration of their complex causes, she then proceeds to address legal and ethical dilemmas such as a patient's refusal of life-saving treatment.

(Issues in Biomedical Ethics)

2005 $74.00

Fictionalism in Metaphysics
Edited by MARK ELI KALDERON

In metaphysics fictionalism is now widely regarded as an option worthy of serious consideration. This volume represents a major benchmark in the debate: it brings together an impressive international team of contributors, whose essays represent the state of the art in various areas of metaphysical controversy, relating to language, mathematics, modality, truth, belief, ontology, and morality.

2005 paper $29.95 cloth $85.00

New in Paperback!
The Oxford Handbook of Metaphysics
Edited by MICHAEL J. LOUX and DEAN W. ZIMMERMAN

Twenty-four of the world's most distinguished specialists provide brand-new essays about "what there is": what kinds of things there are, and what relations hold among entities falling under various categories.

2003 (paper November 2005) paper $45.00 cloth $125 00

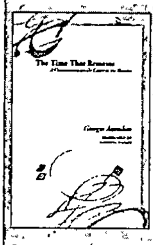
Contextualism in Philosophy
Knowledge, Meaning, and Truth
Edited by GERHARD PREYER AND GEORG PETER

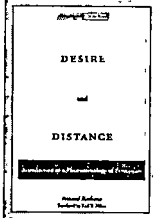
In epistemology and in philosophy of language there is fierce debate about the role of context in knowledge, understanding, and meaning. This volume brings together the debates, in a set of twelve specially written essays representing the latest work by leading figures in the two fields.

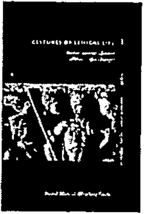
2005 paper $29.95 cloth $99 00

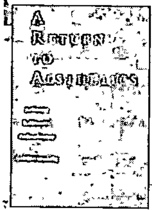
Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS









NEW FROM STANFORD UNIVERSITY PRESS

The Time That Remains

A Commentary on the Letter to the Romans

GIORGIO AGAMBEN

Translated by Patricia Dailey

Agamben seeks to separate the Pauline texts from the history of the Church that canonized them, thus revealing them to be "the fundamental messianic texts of the West." He argues that Paul's Letters are concerned not with the foundation of a new religion but rather with the "messianic" abolition of Jewish law.

$19.95 paper $50.00 cloth

AVAILABLE IN NOVEMBER 2005

Desire and Distance

Introduction to a Phenomenology of Perception

RENAUD BARBARAS

"*Desire and Distance* is based on recent research and presents new ideas on the problem of perception—ideas that are quite enticing. Barbaras is the world's leading Merleau-Ponty scholar, but what makes this book remarkable and philosophically important is that Barbaras distances himself from Merleau-Ponty and develops his own set of concepts with a high level of originality. In my opinion, Barbaras's book is remarkable."

—Leonard Lawlor,
University of Memphis

$22.95 paper $55.00 cloth

Gestures of Ethical Life

Reading Hölderlin's Question of Measure After Heidegger

DAVID MICHAEL KLEINBERG-LEVIN

Carrying forward the problematic of measure and measurelessness that Plato, Aristotle, and Hölderlin posited at the center of their ethics and politics, this book explores ways in which, as the very condition of their freedom, our gestures bear within their most originary sense and sensibility the values, norms, ideals, and prophetic, messianic dreams of a redeemed ethical and political life.

$24.95 paper $65.00 cloth

A Return to Aesthetics

Autonomy, Indifference, and Postmodernism

JONATHAN LOESBERG

A Return to Aesthetics confronts postmodernism's rejection of aesthetics by showing that this critique rests on central concepts of classical aesthetic theory, namely autonomous form, disinterest, and symbolic discourse.

$24.95 paper $65.00 cloth

AVAILABLE IN NOVEMBER 2005

Multiple Arts

The Muses II

JEAN-LUC NANCY,
Edited by SIMON SPARKS

This collection of writings by Jean-Luc Nancy, the renowned French critic and poet, delves into the history of philosophy to locate a fundamentally poetic modus operandi there. The book represents a daring mixture of Nancy's philosophical essays, writings about artworks, and artwork of his own.

$24.95 paper $65.00 cloth

AVAILABLE IN NOVEMBER 2005

Take Care of Freedom and Truth Will Take Care of Itself

Interviews with Richard Rorty

RICHARD RORTY, Edited and with an Introduction by EDUARDO MENDIETA

"Brilliant, lively, ethically inspiring, Richard Rorty's new compilation is an extension of his recent political analysis and diagnosis of the contemporary American condition."

—Drucilla Cornell,
Rutgers University and author of
*Defending Ideals:
War, Democracy and
Political Struggles*

$19.95 paper $50.00 cloth

Stanford
University Press

800.621.2736 www.sup.org

Batta
13.6.06

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NUMBER 11
NOVEMBER 2005

page
547 *Class and Membership* Kit Fine
573 *A Puzzle about Disbelief* Gary Ostertag

BOOK REVIEWS

594 Kurt Raaflaub: *The Discovery of Freedom in Ancient Greece* Zena Hitz
602 NEW BOOKS

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Christopher Peacocke, Philip Pettit, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.
You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2006

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U.S. copyright law. All rights reserved. For more information about our copyright policies, please consult our January issue or contact us.

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc., MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027. Periodicals postage paid at New York, NY, and other mailing offices.
POSTMASTER: Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2005 by the Journal of Philosophy, Inc.

ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY
VOLUME CII, NO. 11, NOVEMBER 2005

CLASS AND MEMBERSHIP*

I wish to describe a construction that is capable of yielding a new solution to the set-theoretic paradoxes. Perhaps what is most distinctive about the construction is the reversal in the roles of the predicate of membership and the ontology of sets. On the usual conception of the cumulative hierarchy of Zermelo-Fraenkel set theory (ZF), we think of the membership predicate as given and of the ontology of sets or classes as something to be made out. Thus given an understanding of membership, we successively carve out the ontology of sets by using the membership predicate to specify which further sets should be added to those that are already taken to exist. Under the present approach, by contrast, we think of the ontology of classes as given and of the membership predicate as something to be made out. Thus given an understanding of the ontology of classes, we successively carve out extensions of the membership predicate by using conditions on the domain of classes to specify which further membership relationships should obtain. What unfolds is not the ontology of sets or classes but the meaning of membership. This "Copernican revolution" in our conception of class membership, once properly implemented, is capable of yielding a theory of classes that is just as natural as the standard theory of ZF and yet far more powerful in the strength of its principles and the scope of its applications.

---

* Much of the material of this paper was presented at a workshop in the philosophy of mathematics at the University of California, Los Angeles, during the academic year 1996–97. It arose from some previous work on Cantorian abstraction and is briefly mentioned in my paper, "Cantorian Abstraction: A Reconstruction and Defense," in this JOURNAL, XCV, 12 (December 1998): 599–634, on pp. 602 and 623. I should like to thank the members of the workshop—including Joseph Almog, David Kaplan, and Tony Martin—for helpful comments. I am also grateful for comments from the audience at talks given at the philosophy department of the University of Melbourne, the Arche Project at St. Andrews, the European Analytic Philosophy Congress at Lisbon, and a seminar at New York University.

0022-362X/05/0211/547–72

I believe that the underlying ideas behind the approach are very intuitive and that they can be understood by someone with little formal background in set theory or logic. For this reason, I have tried to provide an informal exposition of the ideas, one that will be intelligible to a nontechnical reader and yet sufficiently detailed to enable the technical reader to put a precise gloss on what I say. I begin with an informal description of the construction and then present a more precise account of the construction and its principal features (sections I–II); I indicate what kind of set theory it leads to and how the theory may be applied in ways that go far beyond what is possible in ZF (sections III–IV); and I conclude with some philosophical reflections on the resulting conception of class (section V). Results, for the most part, are stated without proof; and it is my hope to provide a fuller and more rigorous exposition elsewhere.[1]

I. THE CONSTRUCTION

We may give an informal description of the construction in terms of an imaginary dialogue between Archangel Gabriel and God. We suppose that God has a full knowledge of the ontology of classes. He is capable of apprehending any class and, in apprehending the class, he is capable of apprehending its members. We may suppose that Gabriel, on the other hand, only has full knowledge of the ontology of sets, that is, of those classes that are small in size relative to the universe as a whole. Like God, he is capable of apprehending any set and, in apprehending the set, he is capable of apprehending its members. But in contrast to God, he lacks full knowledge of the proper classes, that is, of those classes that are not sets. He is capable of apprehending any proper class but, in apprehending the class, he has no awareness of its members; it simply appears to him as an "urelement," an individual without members. We might compare classes and their members to boxes and their contents. All the boxes for God are transparent; in looking at a box, He is aware of its contents. But for Gabriel, it is only the "small" boxes that are transparent. The "big" boxes are opaque and, in looking at them, he has no idea of what their contents might be.

---

[1] There is a loose connection with the theory of Alonzo Church, expounded in "Set Theory with a Universal Set," in Leon Henkin, ed., *Proceedings of the Tarski Symposium*, Symposia in Pure Mathematics, Volume XXV (Providence, RI: American Mathematical Society, 1974), pp. 297–308, and in the Ph.D. thesis of U. Oswald, "Fragmente von 'New Foundations' und Typentheorie" (1976), but they are handicapped by working with an explicit coding for the classes. A related line of investigation has recently been pursued by Øystein Linnebo in "Sets, Properties, and Unrestricted Quantification," to appear in Agustín Rayo and Gabriel Uzquiano, eds., *Absolute Generality* (New York: Oxford, 2006). Perhaps the main differences between my approach and his is that I allow for the classes to be defined by infinitary conditions, for the construction to proceed to a class-inaccessible ordinal, and for the definitions to contain constants for as yet unidentified classes. This results, I believe, in a more powerful and natural theory of classes.

The purpose of the dialogue is to provide Gabriel with the information initially possessed by God. Thus once the dialogue is over, Gabriel's knowledge of the ontology of classes will coincide with God's. The dialogue proceeds in stages; and each stage consists of a number of rounds. At the first stage, Gabriel begins by specifying a condition on classes. Thus he might specify the condition '$x \neq c$', where 'c' is the name of one of the classes that he apprehends or he might specify the condition 'x is finite'. God will then respond by pointing to the corresponding class, that is, to the class whose members are the objects that satisfy the condition. We might imagine that in response to the specification of a condition an hitherto opaque box becomes transparent, or "opens up," and its contents are revealed to be exactly those objects that satisfy the condition.

It is a critical constraint on the conditions that Gabriel may use in specifying a class that they should be intelligible to him in terms of his current understanding of the membership predicate. Thus in so far as he uses the membership predicate, or notions defined in terms of the membership predicate, in specifying a condition, the extension of the predicate should already be apparent to him. What this means in the present case is that in specifying a condition he can only legitimately talk of the members of *sets*, not of classes in general. So, for example, given a particular class c, he can use the condition '$c \in x$' (where 'c' is a name for c) to specify the class of all *sets*, that contain c but he cannot use it to specify the class of all *classes* that contain c since he has as yet no understanding of what it is for c to belong to an arbitrary class. It is this constraint that prevents Gabriel from pulling a "Russell" on God. Gabriel can indeed lay down the condition '$x \notin x$'. But '$\notin$' here must be understood according to his current understanding of '$\in$' as membership in a set. Thus God may appropriately respond by pointing to a proper class that contains all sets that are not members of themselves (and also contains all proper classes). What Gabriel cannot do is to use '$\in$' in such a way as to anticipate what its extension will subsequently become; and so there will be no difficulty in supposing that the condition '$x \notin x$' corresponds to a class as long as that class is not already in the range of the membership predicate by which it is defined.

We may suppose that God and Gabriel go through every condition in turn: Gabriel specifies the condition; and God points to the class that corresponds to the condition (be it a "new" class or one whose identity has already been revealed). The first stage is then complete. There are now two possibilities. One is that the identity of every class has been revealed at some round at this stage. There is then nothing more to be done, since Gabriel has achieved the same understanding of membership as God. The other possibility is that there are classes whose identity has not yet been revealed. They must then move on to the second stage.

This is just like the first stage but with the difference that Gabriel now has an enhanced understanding of membership. For he now knows not only what it is for an object to belong to a set but also what it is for an object to belong to a class of the first stage, that is, to a class that corresponds to a condition that can be specified in terms of set membership. Thus given his enhanced understanding of membership, the old conditions in '$\in$' will have a new meaning and can be used to specify new classes. So, for example, the condition '$\exists y(y \in x)$' will now pick out the nonempty classes that emerged from the first stage and not merely the nonempty sets. Once Gabriel and God have repeated the routine of specifying classes for each condition, the second stage of dialogue will also be complete.

They will then move on to the third stage as long as there are some classes whose identity still needs to be revealed; and, in general, God and Gabriel will move on to a subsequent stage as long as the identity of some classes remains unrevealed. It is possible, of course, that this will force them to go through each finite stage 1, 2, 3,.... They must then proceed to the first transfinite stage ω, at which the extension of the membership predicate will be the "limit" of its extension at the previous stages; to belong to a class at this stage is simply to belong to a class at a previous stage. The dialogue will then continue as before, but with the limit stage as the new starting point. Of course, other limits may be reached in a similar way but eventually, we may suppose, the dialogue will come to an end; the identity of all the classes will have been revealed and Gabriel's understanding of class membership will coincide with God's.

Let us now give more detail. We suppose given the universe $\mathscr{P}$ of all proper classes—which, for formal purposes, we may regard as a collection of urelements (or nonsets). We then take the whole universe $\mathscr{U}$ of *all* classes to consist of all the proper classes in $\mathscr{P}$, all sets of proper classes (including the empty set), all sets whose members are proper classes or sets of proper classes, and so on all the way up through the cumulative hierarchy of sets.[2] In other words, the universe $\mathscr{U}$ is taken to consist of all the sets which can be built up in the standard way from the proper classes, treating these as urelements. Let $\mathscr{S}$ be the property of being a set (which we identify with its extension $\mathscr{U}$-$\mathscr{P}$) and let e_0 be the relation of membership in a set (which we identify with its extension, consisting of the set of ordered pairs (c, d) of members of $\mathscr{U}$ for which $c \in d$). Then the property $\mathscr{S}$ and the relation e_0 correspond to Gabriel's initial understanding of the class-theoretic universe: he knows which of the classes are sets; and he knows which classes belong to which sets.

---

[2] To be exact, we take $\mathscr{P}$ to consist of κ urelements, where κ is a strongly inaccessible cardinal. Let $\mathscr{U}_0 = \mathscr{P}$, $\mathscr{U}_{\alpha+1} = \mathscr{U}_\alpha \cup \wp(\mathscr{U}_\alpha)$, and $\mathscr{U}_\lambda = \cup_{\zeta<\lambda}\mathscr{U}_\zeta$. We then set $\mathscr{U} = \mathscr{U}_\kappa$.

We must now say something about the conditions by which Gabriel may specify a class. Since Gabriel is an angel, we may suppose that he has no difficulty in using infinitary forms of expression—infinitary disjunctions $A_1 \vee A_2 \vee \ldots$, say, or infinitary strings of quantifiers $\exists x_1 \exists x_2 \ldots$. Indeed, we may suppose that he has the same capabilities in regard to the set-theoretic transfinite as we have in regard to the finite. Thus just as we are capable in principle of writing down any finitary disjunction or quantifier string, he is capable of writing down any set-sized disjunction or set-sized quantifier string. We may also suppose, given that he apprehends all the classes, that he is able to provide a name for any given class. Thus the atomic formulas that he may use are those of the form '$s = t$' (s is identical to t); '$S(s)$' (s is a set); and '$s \in t$' (s is a member of t), where s and t are either variables or class names; and the formulas he may use are those that are either atomic or can be formed from the atomic formulas by means of negation, infinitary disjunction, and infinitary quantification.[3]

Given Gabriel's initial understanding of the predicates S and $\in$ (as represented by the model $\mathscr{M}_0 = (\mathscr{U}, \mathscr{S}, e_0))$, each condition $A(x)$ in the single variable x will determine a corresponding extension $|A(x)|_0$, consisting of the classes from $\mathscr{U}$ which satisfy the condition $A(x)$ in the model $\mathscr{M}_0$. Let us call the extension $|A(x)|_0$ *proper* if it is not a set. Then God's responses during the first stage of the dialogue will result in a one-one function δ that maps proper classes of $\mathscr{P}$ onto the family of proper extensions $|A(x)|_0$. Intuitively, $\delta(c) = |A(x)|_0$ if God points to c as the class corresponding to the condition $A(x)$. The function should be one-one since we do not want distinct classes to correspond to coextensive conditions; and the function should be onto since we want there to be a class for any proper extension $|A(x)|_0$ that Gabriel might specify.

Given the assignment function δ, Gabriel may come to a new understanding of $\in$ (as represented by a model $\mathscr{M}_1 = (\mathscr{U}, \mathscr{S}, e_1))$.

---

[3] For certain applications, we may wish to circumscribe Gabriel's means of expression; but present purposes are best served by allowing him to use the full resources of the infinitary language.

According to this new understanding, a class d will belong to a class c just in case it either belongs to c according to the old understanding of $\in$ or is in the extension $|A(x)|_0$ that δ assigns to c (in other words, $e_1 = e_0 \cup \{(d, c): d \in \delta(c)\}$). Let the *extent of a* class c in a model $\mathcal{M} = (\mathcal{U}, \mathcal{S}, e)$ be the collection of its members according to the model (that is, $\{c: (c, d) \in e\}$). Then with c and $A(x)$ understood as above, the extent of c in the new model $\mathcal{M}_1$ will be identical to the extension of $A(x)$ in the old model $\mathcal{M}_0$.

The means by which Gabriel comes to the new understanding of $\in$ on the basis of the old understanding (that is, by which $\mathcal{M}_1$ is formed from $\mathcal{M}_0$) may be applied whatever the initial understanding of $\in$ (or the initial model $\mathcal{M} = (\mathcal{U}, \mathcal{S}, e)$).[4] For we may define the extension $|A(x)|$ of $A(x)$ as before, but now relative to the model $\mathcal{M}$, and we may take an extension $|A(x)|$ to be proper if there is no class in $\mathcal{M}$ whose extent is $|A(x)|$. The assignment δ should now be taken to be a one-one function from a family of proper classes not in range of e (these are proper classes whose identity has not yet been revealed) onto the family of proper extensions; and given δ, the new understanding of $\in$ (or the new model $\mathcal{M}^{\delta} = (\mathcal{U}, \mathcal{S}, e^{\delta})$) may be defined in the same way as before so that $\delta(c)$ is the extent of c. We might call $\mathcal{M}^{\delta}$, in this case, *a predicative expansion of* $\mathcal{M}$.

If all goes well, the interchange between God and Gabriel will result in a sequence of models $\mathcal{M}_0, \mathcal{M}_1, \dots, \mathcal{M}_{\alpha}, \dots, \mathcal{M}_{\lambda}$ with the following features:

> *Start*: The initial model $\mathcal{M}_0$ is the model $(\mathcal{U}, \mathcal{S}, e_0)$, with $\mathcal{U}$ and e_0 understood as before.
> *Successor*: $\mathcal{M}_{\alpha+1}$ is a predicative expansion of $\mathcal{M}_{\alpha}$.
> *Limit*: $\mathcal{M}_{\alpha}$ is the limit of the $\mathcal{M}_{\zeta}$ for $\zeta < \alpha$ when α is a limit ordinal.
> *Finish*: The range of e_{λ} (the membership relation of $\mathcal{M}_{\lambda}$) is $\mathcal{P}$.

Start describes Gabriel's initial understanding of class membership, under which it coincides with membership in a set. Successor and Limit describe the gain in his understanding of class membership: at successor stages, it is given by a predicative expansion; and at limit stages, it is the cumulation of his previous understanding. Finish describes his final understanding of class membership; "all is revealed" and Gabriel's understanding coincides with God's.

---

[4] Assuming that e_0, as defined above, is a subset of e. There is something to be said for taking the whole of $\mathcal{U}$ (and not just $\mathcal{P}$) to be a collection of urelements and allowing e to be an arbitrary relation on $\mathcal{U}$. Starting with e as the empty relation then allows the identity of sets to be revealed in the same way as proper classes.

We might call a sequence of models $\mathcal{M}_0, \mathcal{M}_1,\ldots, \mathcal{M}_\alpha,\ldots, \mathcal{M}_\lambda$ with the above features a *membership* sequence. Its *height* is the size of the universe $\mathcal{U}$ and its *length* is λ. A *transitional* model is one of the models $\mathcal{M}_\alpha$ of a membership sequence $\mathcal{M}_0, \mathcal{M}_1,\ldots,\mathcal{M}_\alpha,\ldots, \mathcal{M}_\lambda$; and a *terminal* model is the final model $\mathcal{M}_\lambda$ of a membership sequence $\mathcal{M}_0, \mathcal{M}_1,\ldots, \mathcal{M}_\alpha,\ldots, \mathcal{M}_\lambda$. The *height* and *length* of $\mathcal{M}_\lambda$ are taken to be the height and length of the membership sequence (it can be shown that membership sequences of different lengths cannot terminate in the same model). It is the terminal membership models that we wish to advocate as models for our theory of classes.

II. MODELS

The construction leaves many questions open; and we must now see how they might be settled.

We should note in the first place that God appears to have a great deal of freedom in how to assign extensions to proper classes. When Gabriel specifies his first condition—the condition $x = x$, say—then God can in principle choose any one of the proper classes in $\mathcal{P}$ as the class that answers to the condition. Usually, nothing of significance will turn on which proper class He chooses. But there is one important exception. Suppose Gabriel specifies the condition '$x \neq c$' and God selects c itself as the corresponding class. Then the selection of c is in a certain sense circular since the class c must be understood by reference to c itself, namely, as the class of classes not identical to c. The selection of classes may also be indirectly circular. Suppose, for example, that d is selected as the class corresponding to the condition '$x \neq c$' and c is selected as the class corresponding to the condition '$x \neq d$'. Then c must be understood in terms of d and d in terms of c. The selection of classes may also be ill founded. Suppose, for example, that c_0 is selected as the class corresponding to the condition '$x \neq c_1$', c_1 as the class corresponding to the condition '$x \neq c_2$', and so on. Then c_0 must be understood in terms of c_1, c_1 in terms of c_2, and so on ad infinitum.

We may wish to avoid irregularities of the sort. To explain what this comes to, we shall need to introduce some further notions, of interest in their own right. We have already come across the notion of what one might call an *extensional* assignment. This is a one-one function δ mapping proper classes from $\mathcal{P}$ into *extensions* (that is, subcollections of the universe $\mathcal{U}$). An *intensional* assignment, on the other hand, is a one-one function Δ that maps classes into *conditions*. We might think of an intensional assignment as a system of definitions of the classes in its domain; where c is such a class, $\Delta(c)$ is its definition. Given an intensional assignment Δ, we may say that the class c *depends upon* or is *defined in terms of* d (relative to Δ) if a constant 'd' for d appears in the condition $\Delta(c)$ by which c is defined. Thus if $\Delta(d)$ is the condition '$x \neq c$' and $\Delta(c)$ is the condition '$x \neq d$', as above, then d will depend upon c and c upon d. We may then say that the intensional assignment Δ is *regular* if there are no classes c_0, c_1, c_2, ... such that c_0 depends upon c_1, c_1 depends upon c_2, and so on ad infinitum. Every definition "bottoms out" in definitions that make no appeal to proper classes.

Suppose that $\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_\alpha, \ldots, \mathcal{M}_\lambda$ is a membership sequence. Then the identity of each class c will first be revealed at some stage α (this is the first ordinal α for which $c \in \mathcal{S} \cup Rg(e_\alpha)$; note that α cannot be a limit ordinal). We call α the *order of* c. So, for example, the order of each set will be 0 and the order of the proper class whose extent is the collection of all sets will be 1. Let Δ be an intensional assignment that is defined on *all* classes. Then Δ will *induce* the membership sequence $\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_\alpha, \ldots, \mathcal{M}_\lambda$ if, for each set c (of order 0), $\Delta(c)$ is the disjunction of the formulas $x = d$ for d a member of c and if, for each proper class c of order $\alpha + 1$, its extent (in $\mathcal{M}_{\alpha+1}$) is identical to the extension $|\Delta(c)|_\alpha$ of the defining condition $\Delta(c)$ for c in the model $\mathcal{M}_\alpha$. In other words, we can suppose that in the dialogue $\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_\alpha, \ldots, \mathcal{M}_\lambda$ the identity of the proper class c is revealed at stage $\alpha + 1$ by means of the condition $\Delta(c)$.

We can now say that a terminal model $\mathcal{M}_\lambda$ is *regular* if it is the terminal model of a membership sequence $\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_\alpha, \ldots, \mathcal{M}_\lambda$ that is induced by a regular intensional assignment. In other words, the model can be seen to result from an interchange between Gabriel and God in which the proper classes are identified by means of a regular or "well-behaved" system of definitions.

It is the regular models, under our approach, which correspond to the well-founded models of ZF. A well-founded model of ZF is one in which there are no regressive *membership* chains $c_0, c_1, c_2, \ldots$, with c_1 a member of c_0, c_2 a member of c_1, and so on ad infinitum. But now what is at issue, under our own conception of class, is that there be no regressive *definitional* chains.

Regular membership sequences have two important properties (which we state without proof):

Automorphism: Suppose $\mathcal{M}_\alpha$ is a regular transitional model and that f is a permutation on the proper classes of order $> \alpha$ (these are the proper classes whose identity has not yet been revealed). Then f may be extended to an automorphism on $\mathcal{M}_\alpha$ (that is, an isomorphism of $\mathcal{M}_\alpha$ onto itself).

Isomorphism: Any two regular terminal models of the same height and length are isomorphic (in other words, if $\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_\alpha, \ldots, \mathcal{M}_\lambda$ and

> $\mathcal{N}_0, \mathcal{N}_1, \ldots, \mathcal{N}_\alpha, \ldots, \mathcal{N}_\lambda$ are two regular membership sequences with universes $\mathcal{U}$ and $\mathcal{V}$ of the same size, then $\mathcal{M}_\lambda$ and $\mathcal{N}_\lambda$ are isomorphic).

According to the first of these results, proper classes whose identities have still not been revealed at a given stage of the dialogue are indistinguishable from one another at that stage. They start off as black boxes and remain so, with no clue as to how they might be distinguished, until the identity of their members has been revealed. This result is far from trivial since, even if we cannot distinguish between two unrevealed classes in terms of their members, we might still be able to distinguish between them in terms of the revealed classes to which they belong.

According to the second result, regular terminal models of the same height and length are essentially unique; their difference lies merely in the identity of the proper classes. Imagine there were a fake God who did not actually know which proper classes, or even which objects, answered to which conditions. Still, he could guess; and as long as his responses were made to accord with a regular system of definitions for the proper classes, he could do just as well as God (assuming, of course, that God's own responses also accord with a regular system of definitions!).

Nothing so far has been said about the existence of membership sequences or membership models. It can in fact be shown that membership sequences of any reasonable length will exist:

> *Existence*: Suppose that the ordinal λ is a nonzero ordinal of cardinality no greater than the height κ of the universe. Then there exists a regular membership sequence $\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_\alpha, \ldots, \mathcal{M}_\lambda$ of length λ.

It can be shown that, in any membership sequence $\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_\alpha, \ldots, \mathcal{M}_\lambda$, the identity of new classes will constantly be revealed (that is, for each ordinal $\alpha+1 < \lambda$, there will be a class of order $\alpha+1$). This means, if κ^+ is the next cardinal after κ, that no membership sequence can be of length at least as great as κ^+, since it would then have to contain more objects than there are in the universe. What this result therefore shows is that the length of a membership sequence can be anything else, from 1 to any ordinal less than κ^+.

Once we fix upon a height κ for the universe, these results then justify us in talking of *the* terminal regular model of given length λ, for $0 < \lambda < \kappa^+$. For Existence will guarantee the existence of such a model; and Isomorphism will guarantee its uniqueness (up to the identity of its proper classes).

But how are we to choose among the terminal models of different length? What might determine whether the dialogue between Gabriel and God ends at one ordinal rather than another? I do not wish to dispute the interest of different choices for the length λ, but there is, I believe, a very natural choice that might be made. For God (given his well-known love of freedom) may let Gabriel decide how far the dialogue is to proceed. He may allow him, at any given stage of the dialogue, to specify a well-ordering; and He, in His turn, will then guarantee to proceed as far as the well-ordering requires. Thus if the well-ordering is of order-type α, He will allow the dialogue to proceed to stage α.

Again, we need some new notions to make this idea precise. Given a universe $\mathscr{U}$ of fixed height κ (which we take to be a strongly inaccessible cardinal), let us say that the ordinal α is *accessible from* the ordinal β if $\alpha > \beta$ and if, in any regular membership sequence $\mathscr{M}_0, \mathscr{M}_1, \ldots, \mathscr{M}_\beta$, there is a formula $A(x, y)$ of two variables x and y whose extension $|A(x, y)|$ in the model $\mathscr{M}_\beta$ is a well-ordering of order-type α. In other words, the subsequent use of α can be anticipated at β. Now let us say that an ordinal is (*class*)-*accessible* if it is accessible from some preceding ordinal. We here use the notion of class–accessibility in rough analogy to the standard notion of accessibility in set theory—which we might, by contrast, call *set*-accessibility.

We are interested in regular membership sequences $\mathscr{M}_0, \mathscr{M}_1, \ldots, \mathscr{M}_\alpha, \ldots, \mathscr{M}_\lambda$ for which λ is class-inaccessible. In such sequences, any ordinal β which is accessible from a nonterminal stage α of the sequence will itself be a nonterminal stage of the sequence. For if β is accessible from α and is either greater than or equal to λ, then it is readily shown that λ is accessible from α and hence not inaccessible after all. Regular membership sequence $\mathscr{M}_0, \mathscr{M}_1, \ldots, \mathscr{M}_\alpha, \ldots, \mathscr{M}_\lambda$ of inaccessible length λ will be in partial conformity with God's policy of leaving the length of the dialogue to Gabriel, since any ordinal stage that Gabriel can anticipate will eventually be used.

We know that any ordinal $\alpha < \kappa$ is accessible. Indeed, at stage 0, Gabriel can find a formula $A(x, y)$ that well-orders all of the ordinals $< \kappa$[5]; and, since its order-type is κ, this shows that κ (and all smaller ordinals) are accessible. We also know that, in any membership sequence $\mathscr{M}_0, \mathscr{M}_1, \ldots, \mathscr{M}_\alpha, \ldots, \mathscr{M}_\lambda$, λ must be less than κ^+. The critical question is therefore whether there is an inaccessible ordinal between κ and κ^+. It is not obvious that there is; for as God complies with the ordinals that Gabriel produces, Gabriel will be in a position to produce even greater ordinals.

---

[5] This is simply the formula $x \in y$ & $\mathrm{Ord}(y)$.

However, it turns out that Gabriel will eventually catch up with himself (the dog will catch his own tail, if I may compare Gabriel to a dog):

Inaccessible Existence: There exists a class-inaccessible ordinal $< \kappa^+$.[6]

By Model Existence, it then follows that there exists a regular membership sequence $\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_\lambda$ and hence a regular terminal model $\mathcal{M}_\lambda$ of inaccessible length λ. We may, in particular, let λ be the first inaccessible ordinal κ^*. The resulting membership sequence $\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_{\kappa^*}$, will then be in *exact* conformity with God's policy of leaving the length of the dialogue to Gabriel. For the nonterminal stages $0, 1, \ldots, \alpha, \ldots$ of the sequence will be exactly the ordinals accessible from below.

I wish to propose the regular terminal models $\mathcal{M}_\lambda$, for λ class-inaccessible, as the intended models for the theory of classes. Just as set-inaccessibility represents a natural closure condition for the formation of sets, so class-inaccessibility represents a natural closure condition for the definition of classes. Of course, this still leaves open what the set-inaccessible or class-inaccessible closure ordinal should be. But one natural choice, in either case, is that the ordinal should be the first of its kind. The intended models of our theory will then be unique up to isomorphism; and indeterminacy in regard to length will be resolved in essentially the same way as indeterminacy in regard to height.[7]

III. AXIOMS

Let us turn to the question of how the resulting theory of classes might be axiomatized. There are different ways in which this might be done but we shall follow the policy of providing an axiomatization that attempts to reflect the actual construction of the models. This has the advantage of making clear how the axioms correspond to the various conditions on the construction even if a conceptually more basic, though less perspicuous, formulation might also be given.

---

[6] In fact, the class-inaccessible ordinals $< \kappa^+$ will be arbitrarily large; for any ordinal $\alpha < \kappa^+$ there will be a class-inaccessible ordinal λ for which $\alpha < \lambda < \kappa^+$.

[7] There are other ways in which the terminal models might be constructed. Thus one might attempt systematically to violate regularity, thereby leading to an anti-foundationalist theory of sets and classes; or one might suppose that the initial model $\mathcal{M}_0$ was endowed with a well-ordering of the universe, thereby allowing a global form of Choice; or one might circumscribe the conditions to reflect various natural constraints on the means of definition.

To this end, we suppose that our terminal models $\mathscr{M}_\lambda$ are endowed with a hierarchy of orders. Thus instead of being of the form $(\mathscr{U}, \mathscr{S}, \epsilon)$, where $\mathscr{S}$ is the collection of sets, they are of the form $(\mathscr{U}, \mathscr{O}, \epsilon)$, where $\mathscr{O}$ is the collection of all orders, that is, all proper classes o_α whose extent is the domain of classes of order α for some ordinal $\alpha < \lambda$. The language of our theory accordingly has three primitive predicates: $=$ for identity; $\in$ for membership; and O for order (O being true of each of the proper classes o_α).[8]

A class x is said to be a *set* (written: $S(x)$) if it belongs to every order, that is, if $\forall y(O(y) \rightarrow x \in y))$. Let us use: r, s, t as variables for sets; o, p, q as variables for orders; and x, y, z as variables for arbitrary classes. Sets are governed by the standard axioms of ZFC:

Null Set: The null class is a set (that is, $\exists y[\forall x(x \notin y)\ \&\ \forall o(y \in o)])$.
Pairing: The pair of any two objects is a set.
(*Set*) *Union*: The union of any set of sets is a set.
Power: The powerset of any set is a set (that is, $\forall s \exists t \forall r(r \subseteq s \rightarrow r \in t))$.
Infinity: There exists a set containing the null set and containing $x \cup \{x\}$ whenever it contains x.
Replacement: $\forall x \exists ! y A(x, y) \rightarrow \forall s \exists t \forall y(y \in t \equiv (\exists x \in s) A(x,y)))$.
Choice: Any set of disjoint nonempty sets enjoys a choice set.[9]
Foundation: Any set of sets has an $\in$-minimal member.
Extensionality: Classes with the same members are the same.

Note that Extensionality has been formulated so as to apply to all classes and not just to sets. Also note that there is no need for the members of a pair to be sets, nor for the members of the image set t in Replacement to be sets.

We have the following three "structural" axioms for the orders:

Containment: Every class belongs to some order $(\forall x \exists o(x \in o))$.
Well-foundedness: The orders are well-founded $(\forall x[\exists y(y \in x)\ \&\ \forall y(y \in x \rightarrow O(y))] \rightarrow \bigcap x \in x)$.
Embedding: Any well-ordered class is embeddable in the orders $(\forall R(R$ is a well-ordering $\rightarrow \exists \bar{f}(\text{Func}(\bar{f})\ \&\ \text{Dom}(\bar{f}) = \text{Field}(R)\ \&\ \forall y(y \in \text{Range}(f) \rightarrow O(y))\ \&\ \forall x_1 \forall x_2(x_1 R x_2 \rightarrow f(x_1) \subset f(x_2)))))$.

---

[8] Under an alternative approach, we would define the notion of order: a *set* could be taken to be a class whose Frege-Russell cardinal $\{y: y \sim x\}$ exists; and an *order* could then be taken to be a class that belongs to the range of some function f with domain $\alpha > 0$ for which $f(0) =$ the class of all sets, $f(\beta + 1)$ is the class of classes defined on $f(\beta)$ (in a sense to be explained), and $f(\lambda) = \cup\{f(\beta): \beta < \lambda\}$ for λ a limit ordinal. The axioms concerning order would then be derived from more fundamental axioms using the definitions.

[9] In the absence of some other axioms, we may also require an axiom scheme of dependent choice: $\forall x \exists y A(x, y) \rightarrow \exists s \forall x \in s \exists y \in s A(x,y)$.

From Well-foundedness, it readily follows that the orders are connected ($\forall o \forall p(o \subseteq p \vee p \subseteq o)$). The last of these axioms states that no well-ordering will take us "out" of the orders. With the help of this axiom, it may be shown that there is no last order ($\forall o \exists p(o \subset p)$). We use o_0 for the first order, containing all sets and o^+ for the next order after o, that is, for the first order p for which $o \subset p$.

We next have some axioms governing the "content" of the orders. Most significant of these is the axiom of class comprehension. What this tells us, roughly speaking, is that any class definable in terms of membership in a class of order o will belong to the next order o^+. However, its proper formulation calls for some care. In defining a new class of order o^+, o-membership, that is, membership in a class of o, may be presupposed. This may be represented by a formula of the form $x \in y$ & $y \in o$. The property of being an order may also be presupposed as long as the order in question belongs to the given order o. This property, thus qualified, may be represented by a formula of the form $O(x)$ & $x \in o$.[10] But we may also presuppose membership in any class from a set of classes in o^+. This is because any such class will have a definition in terms of o-membership; and so instead of talking of membership in the class we can talk of satisfaction of the condition by which it is defined. For example, $\exists y(x \in y \;\&\; y \in s)$ may be replaced by $A_1(x) \vee A_2(x) \vee \ldots$, where $A_1(x)$, $A_2(x)$, ... run through all the defining conditions of the classes in s.

Given a formula $A(x)$ (perhaps containing other free variables besides x) and a term t, let $A^t(x)$ be the result of replacing each occurrence of a subformula $y \in z$ with $y \in z \;\&\; z \in t$ and each occurrence of a subformula $O(y)$ with $O(y) \;\&\; y \in t$. We then have:

Class Comprehension: $s \subseteq o^+ \rightarrow \exists y(y \in o^+ \;\&\; \forall x(x \in y \leftrightarrow A^{s \cup o}(x)))$.

This axiom is a kind of analogue of the Separation axiom in ZF. But in keeping with the "Copernican revolution," the restriction on the defining condition $A(x)$ is not now imposed on the universe under consideration but on the membership relation.

The comprehension axiom may be broken down into a finite number of axioms that reflect the construction of the formulas by which the given class is defined (in much the same way as in the finite axiomatization of von Neumann-Bernays-Gödel set theory (NBG)). Given a set I, we take an *I-relation* to be a class of functions whose domain is I (thinking of the members of the index set I as the argument-places of the relation). We may then use the following finite set of individual axioms in place of Class Comprehension:

---

[10] To justify this use of O in terms of the original construction we must show that the collection of all orders o_β, $\beta < \alpha$, is definable at stage α.

Identity: The identity relation belongs to o^+ ($\{f\colon \mathrm{Dom}(f) = \{i, j\} \,\&\, f(i) = f(j)\} \in o^+$).

Constancy: The constant relation belongs to o^+ ($\{f\colon \mathrm{Dom}(f) = \{i\} \,\&\, f(i) = x\} \in o^+$).

Membership: The membership relation for o belongs to o^+ ($\{f\colon \mathrm{Dom}(f) = \{i, j\} \,\&\, f(i) \in f(j) \,\&\, f(j) \in o\} \in o^+$).

Negation: If the I-relation R belongs to o^+ then so does its negation $\sim R$ ($R \in o^+$ & R is an I-relation $\rightarrow \{f\colon \mathrm{Dom}(f) = I \,\&\, f \notin R\} \in o^+$).

Disjunction: If the set s of I-relations belongs to o^+, then so does its disjunction $\vee s$ ($\forall x \in s$(x is an I-relation) $\rightarrow \bigcup s \in o^+$).

Padding: If the I-relation R belongs to o^+, then so does the J-expansion of R for J a superset of I ($R \in o^+ \,\&\, \mathrm{Dom}(R) = I \,\&\, I \subseteq J \rightarrow \{f\colon \mathrm{Dom}(f) = J \,\&\, (\exists g \in R)(g \subseteq f)\} \in o^+$).

Projection: If the I-relation R belongs to o^+, then so does the J-restriction of R for J a subset of I (($R \in o^+ \,\&\, \mathrm{Dom}(R) = I \,\&\, J \subseteq I \rightarrow \{f\colon \mathrm{Dom}(f) = J \,\&\, (\exists g \in R)(f \subseteq g)\} \in o^+$).

Class: If the $\{i\}$-relation R belongs to o^+, then so does the corresponding class ($R \in o^+ \,\&\, \mathrm{Dom}(R) = \{i\} \rightarrow \{f(i)\colon f \in R\} \in o+$).

We might say that a class y is *o-closed* if it satisfies the eight conditions above on o^+ and that x is *definable* on o if it belongs to every o-closure. Comprehension says, in effect, that o^+ is o-closed. But according to the construction, it is the smallest o-closure. We may therefore also lay down:

Converse Comprehension: Any class of order $o \supset o_0$ is definable on some preceding order $p \subset o$.

Finally, we need some axioms to reflect the regularity of the construction. The notion of dependence previously defined was relative to a system of definitions. We are interested in making use of a corresponding absolute notion. To this end, we define $x \prec y$ by $\exists o^+(y \in (o^+\text{-}o) \,\&\, x \in \bar{o} \,\&\, (S(\bar{o} \cap y) \rightarrow x \in y) \,\&\, (S(\bar{o} \cap \bar{y}) \rightarrow x \in \bar{y}))$ ($\bar{t}$ here is the absolute complement of t).[11] What justifies this definition is that in any regular model, the proper class y can be defined in terms of the orders preceding o and a set of constants from $\bar{o}$, where these can be taken to belong to y if $\bar{o} \cap y$ is a set and can be taken to belong to $\bar{y}$ if $\bar{o} \cap \bar{y}$ is a set. The orders preceding o give rise to no "irregularities" and can therefore be ignored. We therefore require:

---

[11] Alternatively, we could use the notion *of defining from constants*, as explained below, and define $x \prec y$ by $\exists o^+(y \in (o^+ - o) \,\&\, x \in \,\&\, \forall z(\forall u(u \in z \rightarrow u \in o \vee O(o) \,\&\, u \in o) \,\&\,$ (y is definable from o using only z) $\rightarrow x \in z$).

> *Class Dependence*: $\prec$ is well founded, that is, any nonempty class contains a $\prec$-minimal member.

We might also wish to state that every class has a "regular" definition. Say that R is an *o-definable relation using only z as constants* if it is an I-relation for some set I and if it belongs to every class satisfying the conditions above with the exception of Class and with the application of Constant restricted to classes $x \in z$; and say that y is a *definable on o using only z as constants* if, for some o-definable $\{i\}$-relation R using only z as constants, $y = \{f(i) : f \in R\}$. We then have the following strengthening of Converse Comprehension.

> *Class Regularity*: Any class y of order $o \supset o_0$ is definable on some preceding order $p \subset o$ using only $\{x\colon x \prec y\} \cup \{q\colon q \subset p\}$ as constants.

I am not here concerned with the deductive development of the theory, although it is worth noting that we cannot in general make use of Separation or Replacement. Where these axioms would be used in ZF, we must instead use Embedding to find a reasonably rich membership-theoretic environment (that is, a sufficiently large order) in which the required definitions can be made.

However, it is worth noting one striking consequence of the axioms:

> *Class Union*: The union $\{z\colon \exists y(z \in y \mathbin{\&} y \in x)\}$ of any class x exists.

This result is by no means obvious. For even though the class x is revealed at some stage, there is no reason to suppose that there is some stage at which all its members are revealed and so no reason to suppose that there is some stage at which the formula $\exists y(z \in y \mathbin{\&} y \in x)$ will pick out the union of x. However, we may draw upon the peculiarities of regular models to show that this circumstance is no impediment to defining the union. Let me give a sketch of the proof.

Suppose first that x is a set. Then each member y of x belongs to some order and hence to a first order $o(y)$. So by Replacement, there exists the set $\{o(y) : y \in x\}$ and so, by Set Union, its union $o = \cup\{o(y) : y \in x\}$ also exists. It can be shown that o is also an order, of which x is both a member (since it is a set) and a subset. But then by Class Comprehension, there will exist a class $\{z\colon \exists y(z \in y \mathbin{\&} y \in o \mathbin{\&} y \in x \mathbin{\&} x \in o)\}$, which can be shown to be identical to $\{z\colon \exists y(z \in y \mathbin{\&} y \in x)\}$.

Suppose now that x is a proper class (that is, not a set). It can then be shown that $x \in o^+ - o$ for some order o. Now $x = (o \cap x) \cup (\bar{o} \cap x)$ and so $\cup\, x = \cup(o \cap x) \cup \cup\, (\bar{o} \cap x)$ will exist as long as both $\cup(o \cap x)$ and $\cup(\bar{o} \cap x)$ exist. But $\cup(o \cap x) = \{z : \exists y(z \in y \mathbin{\&} y \in o \mathbin{\&} y \in x)\}$, which exists by Class Comprehension (given that $x \in o^{+}$). So it remains to show that $\bigcup(\bar{o} \cap x)$ exists. Now $\bar{o} \cap x$ or $\bar{o} \cap \bar{x}$ is a set (since otherwise $x \prec x$, contrary to Dependence). If $\bar{o} \cap x$ is set, then $\bigcup(\bar{o} \cap x)$ exists by the first part. So suppose $\bar{o} \cap \bar{x}$ is a set. It can then be shown that $\bar{o} \cap x$ must have as a member some class and its complement. For it can be shown that $\bar{o}$ is not a set and that it is closed under complements. But if $\bar{o} \cap x$ did not contain some class and its complement, then $\bar{o} \cap \bar{x}$ must contain any class of $\bar{o}$ or its complement. So mapping each member of $\bar{o} \cap \bar{x}$ into its complement, we can use Replacement to show that the set x' of complements of the members of $\bar{o} \cap \bar{x}$ is a set. Hence $\bar{o} \cap \bar{x} \cup x' = \bar{o}$ and is a set, contrary to the fact that $\bar{o}$ is not a set. But given that $\bar{o} \cap x$ contains a class and its complement, $\bigcup(\bar{o} \cap x)$ will be the universal class and will therefore exist.

IV. APPLICATIONS

Let me briefly indicate some of the ways in which the resulting theory of classes can be applied and how, in particular, it might be more flexible or powerful in its applications than ZF.

ZFC. We should first note that ZFC sits inside our theory. For let us define a class x to be a *hereditary set* if it is a set, if its members are sets, if the members of its members are sets, and so on.[12]

The axioms of ZFC can then be shown to hold for the hereditary sets. If, for example, x is an hereditary set then its powerset will exist by Power; and it is then readily shown to be an hereditary set. Thus anything that can be done within ZFC can be done within our theory under a suitable restriction of the quantifiers.

Boolean Algebra. ZFC is not compatible with Boolean Algebra, that is, with the existence of complements and arbitrary unions. For if the complement $\bar{x}$ of some set x were to exist then so would their union $x \cup \bar{x}$, which is the universal class V. But Separation could then be applied to V to give a version of Russell's paradox. Our theory, on the other hand, implies that the domain of classes forms a Boolean algebra. For the existence of complements follows directly from Class Comprehension and, as we have seen, the existence of arbitrary unions can be derived from the other axioms.

There are many applications in which it is desirable that the domain of classes should form a Boolean algebra. We may wish, for example, to give a semantics for a language in which the quantifiers range over all objects. The extension of a negative predicate $\sim P$ should then be taken to be the complement of the extension of the negated predicate P and, likewise, the extension of an existential predicate $\lambda x \exists y Rxy$ should be taken to be the union of the extensions of the instancing predicates $\lambda x Rxa$, where a is the name of an arbitrary object from the domain. Or again, the natural definition of $\cap$ requires that the union $\cap\varnothing$ of the empty set $\varnothing$ should itself be the empty set (since there is no object that belongs to some member of $\varnothing$). Similarly, the natural definition of $\cap$ requires that the intersection $\cap\varnothing$ of the empty set should be the universal class (since any object belongs to all members of $\varnothing$). But this natural definition of $\cap$ and the resulting duality with $\cup$ is impossible without the universal class.

---

[12] The transitive closure of a class x is the smallest class y for which $x \subseteq y \;\&\; \forall u \forall v(u \in v \;\&\; v \in y \rightarrow u \in y)$. A hereditary set x is then one whose transitive closure consists entirely of sets.

Cardinals, Order-types, and Abstractions. Frege and Russell defined the cardinal card(x) of a set x to be in effect the class of all y equinumerous with x ($\{y: y \sim x\}$). But this natural definition of cardinal is unavailable in ZF. Indeed, it can be shown that, with the single exception of the null set, the Frege-Russell cardinal of a set will never exist.

There is no such difficulty within our class theory. Indeed, it can be shown that the cardinality of any set x will belong to o_1. For it will follow from Class Comprehension that $\{y: y \sim x\}$ will belong to i_o as long as the condition $y \sim x$ of there being a 1-1 correspondence between x and y is formulated with the membership predicate restricted to sets. But this is equivalent to the condition $y \sim x$ formulated without any such restriction since a one-to-one correspondence f between the set x and some class y must always be a set. Moreover, the cardinality function itself, taking each set x into card(x), will belong to o_2 since it may be defined over o_1, as the class of pairs (x, y) for which x is a set and y is the cardinality of x.

Similarly for the order-type of a set ordering, the isomorphism-type of a set structure and many other instances of definition "by abstraction." The greater power of class theory in this regard is attributable to its greater power in establishing the existence of equivalence relations. Suppose that R is an equivalence relation (considered as a class of ordered pairs). Then we may establish in the usual way that there exists a function f such that, for every x and y in the field of R, xRy iff $f(x) = f(y)$. This result is common to both standard set theory and our class theory. But in class theory, we may establish the existence of a much broader range of equivalence relations. For suppose $A(x,y)$ is a formula, D a class, and o an order for which $(\forall x,y{\in}D)(A(x, y) \leftrightarrow A^o(x,y))$ (the restriction of membership to o making no difference to the application of $A(x, y)$ to the members of D). It then follows from Class Comprehension that there will exist a relation R for which

$(\forall x,y \in D)(xRy \leftrightarrow A^o(x, y))$ and hence for which $(\forall x,y \in D)(xRy \leftrightarrow A(x,y))$. Thus in the case of cardinality, we may take D to be the class of all sets, o to be o_0, and $A(x, y)$ to be $x \sim y$.

In other applications we may want D to include proper classes and we may also want o to be an order superseding o_0. Suppose, for example, that we want to define the natural numbers modulo n. Then the condition $A(x, y)$ of having the same remainder upon division by n will apply to the domain D of natural numbers which, under the Frege-Russell treatment, will be proper classes. Or, again, suppose we wish to abstract on infinitary conditions which pick out the same extension on the order o. We then need to restrict the condition $A(x, y)$ of picking out the same extension to the succeeding order o^+.

However, we are still unable to extend the definition of the Frege-Russell cardinal to proper classes. The difficulty is that to understand $\{y: y \sim x\}$ we need to understand what it is for an object to belong to an arbitrary class of the same size as x; and such an understanding, when x is a proper class, would never appear to be available to us, since proper classes of the same size as x are "always in the making." Indeed, it may actually be established that $\{y: y = x\}$ never exists when x is a proper class. Thus even though we are in a position to talk of Frege-Russell cardinals in place of the cardinals of ZF, we are still not in a position, given the new classes available to us, to talk of the Frege-Russell cardinal of an arbitrary class.[13]

Extending the Ordinals. We here understand an ordinal in the sense of von Neumann, in which it is identified with the class of its predecessors rather than with a class of similar well-orderings. Thus $0 = 0$, $1 = \{0\}$, $2 = \{0, 1\}$, ... , $\omega = (0, 1, 2, \ldots\}$, and so on. The von Neumann ordinals of our class theory extend far beyond those of ZF, that is, the set ordinals. For Class Comprehension gives us that the class ∞ of all set ordinals, that is, the first ordinal greater than each set ordinal, will belong to o_1. The ordinals $\infty + 1$ $(= \infty \cup \{\infty\})$, $\infty + 2$ $(= \infty \cup \{\infty \cup \{\infty\}\})$, and, in general, $\infty + \alpha$, for α a set ordinal, will also belong to o_1, since they may be defined by explicit enumeration given the definition of ∞. Similarly, the class ∞_o of all ordinals of order o will be a new ordinal of o^+, as will the ordinals of the form $\infty_o + \alpha$, for α a set ordinal.

These new ordinals are very useful in developing the theory of classes. Thus it may be shown that there is a one-one correspondence o_α between ordinals and orders with the property that $\alpha < \beta$ iff $o_\alpha \subset$

---

[13] Though, by an analogue of the "Scott trick," we are able to obtain a shrunken version of the Frege-Russell cardinal for each class x.

o_β. Each order is given an ordinal "level"; and each ordinal is the level of some order. Or again, we may develop an account of definition by transfinite induction that takes us beyond the ordinals $< \infty$. We may then use the ordinals to define a very natural hierarchy of classical truth predicates T_α, one for each ordinal α, where T_0 will apply to all the truths of set theory and where T_α will apply to the truths of a language containing each of the truth predicates T_β, for $\beta < \alpha$. In this hierarchy, we are able to make the truth-theoretic jump to the next higher level but without making the usual set-theoretic jump to larger and larger inaccessible cardinals.

The new ordinals are also useful even if our primary interest is in the set ordinals. The theory of ordinals within ZF suffers from some unfortunate anomalies. In the case of definition by transfinite induction, it requires a separate treatment for the case in which a function is defined on the domain of a given ordinal and in the case in which a function is defined on the domain of all ordinals. One would like to have a uniform treatment of both cases; and one would also prefer, in the second case, to have a single general statement of the principle of definition by transfinite induction rather than a scheme in which the relevant instances of the principle are given by individual formulas of the theory. To some extent, this anomaly is removed in a system like NBG which admits proper classes, since then the principle of definition by transfinite induction can be stated for all class-ordinals, including the proper class ∞ of all set ordinals. But certain highly general desirable principles of the theory of ordinals are then lost. Thus even though it is true in ZF that each well-ordered set has a (set) ordinal, it will not be true in NBG that each well-ordered class has a (class) ordinal, for we can find a well-ordered proper class whose ordinal should be $\infty + 1$, and yet $\infty + 1$ cannot exist since it would then have to have ∞, which is a proper class, as one of its members.

These difficulties are considerably alleviated within our own class theory. For we can provide a general statement of the principle of definition by transfinite induction, one with application both to the domain of a set ordinal and the domain of all set ordinals, and yet retain the principle that every well-ordered class has an ordinal (and other desirable principles of this sort). It has to be admitted, though, that some anomalies remain. For one can provide a schematic statement of definition by transfinite induction, one applicable to the domain of all ordinals, within our theory as well. However, the proof of the schematic principle is rather different from the proof in the case in which the domain of definition is given by an ordinal; and so it is not clear that there should be a general principle that covers both cases, even if one would prefer to have a single general statement in place of a scheme.[14]

Truth and Validity. It has often been remarked that if we are to do model theory within ZF, then the domain of a model must be taken to be a set. But this means that any interpretation of a language whose domain is not a set will fail to correspond to a model and this will apply, in particular, to any interpretation whose domain is the same as the intended domain of the theory, ZF, in which the models are being discussed. If we define validity as truth in all models, then the notion of validity will suffer from a corresponding limitation, since there will be potentially falsifying interpretations that will fail to be represented by any model.

Again, these difficulties are considerably alleviated within our class theory. The theory admits proper classes, and the universal class in particular; and there is no impediment to defining a model with such a class as domain or in defining what it for a sentence to be true in such a model (or for a formula to be satisfied). The range of unrepresented interpretations is therefore considerably reduced. But not altogether. For the intended model of the theory itself will not constitute a class. Its domain (the universal class) will indeed exist but not the extension of $\in$. In the light of Tarski's theorem on truth, this result is unavoidable since otherwise the theory would contain its own truth predicate. What we may say is that, given this unavoidable limitation, our class theory provides a much more adequate account of truth-in-a-model and of validity than the standard account. For we can countenance all of the models countenanced by ZF and many more besides. Instead of adopting a drastic restriction on the ontology of the models under considerations, we are able to adopt a much more modest restriction on their "ideology" or set of relations. Large models can be countenanced, and even large models with a highly complex ideology, as long as their ideology can be confined to a given order.

We might define validity as truth in all models whose domain is the whole universe (and not merely as truth in all models whose domain is a subclass of the universe). We would then appear to get the logic of Harvey Friedman's "theory of everything."[15] So, for example, Friedman takes the formula $\exists x \exists y(x \neq y \,\&\, (Rxy \leftrightarrow Ryx))$ to be valid (whatever the relation R it will fail to distinguish between two objects). But this will be so under the proposed definition of validity since, if R is taken to be a class of ordered pairs, then there will be a proper class of objects, out of its "range" so to speak, between which it will be unable to distinguish. Thus the present theory of classes provides a natural setting, and some kind of vindication, for Friedman's view.

---

[14] The reader should bear in mind that we cannot in general appeal to Replacement in establishing principles of definition by transfinite induction and so the statement and proofs of the principle will go somewhat differently than in the case of ZF.

[15] In "A Complete Theory of Everything" (1999) at www.math.ohio-state.edu/~friedman.

Hybrid Theories. A rather different kind of application concerns the case in which set theory is combined with other theories which appear to call for an unrestricted principle of comprehension. One example concerns the combination of set theory with mereology, the theory of part-whole. It is natural to assume, within mereology, that for any condition $A(x)$ that is satisfied by at least one object there exists the sum of objects satisfying the condition. This condition, in the context of mereology, does not lead to contradiction but it does when combined with set theory and certain very modest assumptions concerning the mereological structure of sets.[16] And similar difficulties can arise when one attempts to combine set theory with the theory of properties or propositions or with a general theory of objects or with various theories of abstraction in the style of Frege.

It is not clear what to say in response to such difficulties; for one does not want to give up the principles of the separate theories and nor does one want to give up the modest principles by which they are linked. One possible line of solution is to state the comprehension principle for sums (or for the entities in question) with respect to classes rather than conditions. Thus it will be maintained that for any class there will exist a sum of the objects belonging to the class. Given our underlying theory of classes, there will then exist many of the sums that one wants to exist—the sum of all objects, for example, or the sum of all urelements—and yet contradiction will be avoided.

V. PHILOSOPHICAL REFLECTIONS

I should like to make some general remarks about how our theory relates to various different views on the nature of classes. Current thinking about set theory has been informed by two main conceptions of class—the combinatorial and the logical. According to the combinatorial conception, a class is taken to derive from its members.

---

[16] Namely that a singleton is part of a sum of singletons only if its one of those singletons. Essentially the only assumptions we need from set theory and mereology are the existence of singletons and the existence of sums. The present proposal is briefly mentioned towards the end of my reply to Shapiro in "Replies to Comments on 'Limits of Abstraction'," *Philosophical Studies*, CXX, 3 (2005): 367–95. See Uzquiano's paper, "The Price of Universality," to appear in Rayo and Uzquiano, eds., *Absolute Generality*, for a recent discussion of the difficulty.

We might say, according to this conception, that there should be an *account* or *definition* of what any class is in terms of its members. Thus if the members of the class c are $d_1, d_2, \ldots$, then there should be an account of what c is in terms of $d_1, d_2, \ldots$ (and in terms of these objects alone).[17]

According to the logical conception, on the other hand, a class is taken to derive from a corresponding condition. There will be an account or definition of any class as the class of objects corresponding to such and such a condition. Thus if the members of the class c are $d_1, d_2, \ldots$, then there will be an account of what c is in terms of a suitable condition $A(x)$, as long as $d_v, d_2, \ldots$ are indeed the objects that meet the condition.[18]

It is a natural requirement on any system of definitions that it should be noncircular or, more generally, well founded. In other words, it should never turn out that c_0 is defined with the help of c_1, c_1 with the help of c_2, and so on ad infinitum. Any definition, once it is unraveled, should bottom out in objects that are either not to be defined or are to be defined with the help of no other objects. This is a fundamental respect, though not the only one, in which definitions differ from mere descriptions.

The meaning of this requirement in the case of the combinatorial conception of class is clear. It means that it should never turn out that c_1 is a member of c_0 c_2 a member of c_1, and so on ad infinitum. Any backward membership chain should terminate with urelements or the null class, as with the standard axiom of foundation for ZF with urelements.[19]

---

[17] We may leave open how exactly the account should go. Thus one might suppose that there is a primitive operation { } for forming sets from objects and that c is the application of this operation to $d_1, d_2, \ldots$, or one might suppose that c is to be understood as the *class* whose *members* are $d_1, d_2, \ldots$. Of course, an important difference between the combinatorial and logical definitions is that the combinatorial definitions are effectively unique while the logical definitions are not. I should note that these remarks on the nature of a class might be placed within the general essentialist framework of my paper, "Essence and Modality," *Philosophical Perspectives*, VIII (1994): 1–16, although this will not be required for an understanding of what I say.

[18] We might characterize the connection between the class c and a defining condition $A(x)$ in terms of membership. But a class, on the logical approach, is more plausibly taken to be the abstract common to all co-extensive conditions—as with Frege's original conception of an extension or course-of-values. Thanks to Carrie Jenkins for helping me get clear on this.

[19] Actually, this is to oversimplify. For the class c_0 might have been defined with the help of the urelement c_1, and the urelement c_1, in its turn, with the help of c_0. Such definitional cycles should also be excluded. The present considerations, for this reason, are best seen as relating to classes whose definition does not require the appeal to urelements or, at least, to urelements that must themselves be understood in terms of classes.

The meaning of this requirement in the case of the logical conception of class is not so clear, however. For how is the condition by which the class is given to be understood? What is required, if I may put it this way, to give an account of the condition itself? This raises several issues but there are two which are especially significant in the present context. The first concerns the range of the quantifiers (or variables) by which the condition is specified. It might be thought that an understanding of a given range of quantification already presupposes an understanding of each of the objects in the range of the quantifier. We then get a form of the famous vicious circle principle of Russell, according to which no class can be defined in terms of a condition whose variables range over the very class in question.

The other issue concerns the combinatorial complexity of the condition itself. Suppose the domain of natural numbers is somehow given to us. Then can I define a class of natural numbers by an infinitary condition such as 'x is prime $\vee$ x is the square of a prime $\vee$ x is the cube of a prime $\vee$ ...'? It might be thought that, in so far as such an infinitary condition is intelligible, we must possess some finitary means of picking out its disjuncts.[20] But if this is so, then the class might be directly specified by these means—as the numbers that are some power of a prime, say, or as the numbers that satisfy a condition of the form 'x is the n-th power of a prime'. Thus the infinitary conditions themselves could not be used in defining a class or, at least, their use should always be dispensable in favor of finitary conditions.

This combination of positions—that the defining conditions should be finitary and subject to the vicious circle principle—yields the standard predicative view of classes and, indeed, this view has constituted the principal way of developing the logical conception. Part of what makes the view so attractive is a feature it shares with the combinatorial conception, namely, that we might regard the definition of a class as a means for *introducing* the class into the ontology. Prior to giving the definition, there is no need to presuppose that the class exists; and once the definition has been made, we then have a basis for supposing that it does exist. Thus we might think of the combinatorial approach as introducing a new class by selecting its members from a pre-existing ontology and we might think of the logical conception as introducing a new class by specifying a condition whose manner of application is to be understood in terms of the pre-existing ontology.

---

[20] Or, objectively put, any account of what the condition is must involve a finitary specification of its disjuncts. I shall sometimes talk something subjectively in this way when I have something objective in mind.

Indeed, it seems clear that if a defining condition is to be regarded as *introducing* a class into the ontology, then the vicious circle principle should be satisfied, that is, the range of the variables by which the condition is specified should not include the very class in question. But it might also be argued, once we adopt the vicious circle principle, that the defining conditions should be finitary—or, at least if we are to have a position that is distinct from the combinatorial conception. For suppose infinitary means of expression were allowed. Then what is to stop the predicativist, whenever the combinatorialist defines a set c in terms of its members d_1, $d_2, \ldots$, to define the set by means of the condition $x = d_1 \vee x = d_2 \vee \ldots$? There would appear to be no reasonable basis for drawing a line between the sets that are so specifiable and those that are not.[21]

However, this picture of classes as being *introduced* by their defining condition is not built into the idea of a defining condition as such. The defining conditions say what the class is but it is no part of their remit, so to say, to indicate how the class got into the ontology in the first place. We therefore have the possibility of an alternative picture, one in which the defining conditions serve to *reveal* the identity of the classes rather than to *introduce* them into the ontology. The classes are already there, so to speak, and the purpose of the successive definitions of the classes is not to make clear how they got there but to make clear what they are.

On this alternative picture, what a condition should properly be taken to presuppose (and what may be seen as blocking the paradoxes) is not the objects in the range of its variables but the extension of its membership predicate. For, as a rule, the condition will be stated by means of a membership predicate, relating a class to its members. But we cannot be taken to understand what it is for an object to belong to a given class unless we already understand what the class is. Thus what takes the place of the vicious circle principle is the principle that a class should not be defined by means of a condition whose membership predicate relates that class to its members. Of course, this constraint already follows from the vicious circle principle, for if we cannot talk of the class then we certainly cannot talk of its members. But it is a much weaker constraint, since we can still talk about the class in the same way that we might talk about an urelement—as something to be quantified over, or to which something might be identical, or as a member of a class. We can presuppose its existence, and even its objecthood, but not its existence *as* a class.

---

[21] This is not, of course, to deny the technical possibility of adopting weaker infinitary languages that might only allow us to specify some of the sets.

On this alternative picture, we also have a rationale for dropping the requirement that the defining conditions should be finitary. For we may simply require that the conditions be set-sized, that is, that they should be of no greater combinatorial complexity than the sets themselves. Thus whatever it is that is taken to prevent the sets from being larger than they are can also be taken to prevent the conditions from being larger than they are. This liberalized constraint did not result in a distinctive position when the vicious circle principle was in effect, since the domain of pre-existing objects to which a defining condition applied was always a set and so the subclasses of the domain that could be defined by logical means was always identical to those that could be defined by combinatorial means. But now the defining conditions are taken to apply to the whole universe of classes, which is a proper class, and so there is a possibility that some classes definable by logical means will not be definable by combinatorial means. Indeed, the universal class is such a class for it is definable by the condition '$x = x$' and yet, given that it is ill founded, there is no way it can be taken to exist under the combinatorial conception. It is of course still true, given the liberalized constraint on conditions, that every class definable by combinatorial means will be definable by logical means; and so the present viewpoint will result in an *expansion* of the classes that are taken to exist under the combinatorial approach and not in a restriction, as with the standard predicative approach.[22]

We thereby obtain a distinctively platonistic gloss on the logical conception of class. The general idea behind the logical conception is that any class should be understood by means of a defining condition. But depending upon one's general philosophical position, one can have very different views as to what the defining conditions might properly be taken to be. For the constructivist or nominalist, they should be finitary and they should probably also conform to the vicious circle principle. But there would appear to be no good reason for the platonist to accept either of these constraints; the classes, for him, are just there and it is hard to see why he should accept any finitary limitations on how they might be defined. Indeed, the adoption of the logical conception helps to make clear how thoroughly platonistic his position is; for he can no longer think of the ontology of classes as built up from its members, as with the standard iterative conception, since the whole ontology must itself be presupposed in order to account for the identity of particular members.[23]

---

[22] One might adopt our substitute for the vicious circle principle and yet retain the restriction to finitary means of expression. One would then obtain a hybrid position that was narrower than the combinatorial position in certain respects and yet broader in others.

I conclude with a question about which I find it hard to have a firm opinion. Should we think of sets as a special kind of class, differing merely in their size, or should we think of sets and classes as distinct kinds of object? It is clearly preferable, on general theoretical grounds, to make do with one kind of object; and this had better be *class* since the classes outstrip the sets. If each set is a class, then it may be taken to be the class defined by the enumerative condition $x = c_1 \vee x = c_2 \vee \ldots$, where $c_1, c_2, \ldots$ are the members of the set. This condition will, of course, in general be infinitary. The critical question now is whether in providing this account of the set we are presupposing the existence of the condition and hence the existence of something that is a set or set-like (since the disjunction must "collect together" its disjuncts). Now when the condition is finitary it would not be thought that its existence was presupposed. In giving an account of the null class as the class without any members, for example, I am not presupposing the condition of being nonself-identical or of having no members. But it might be thought that infinitary conditions are different since there is no way of getting at them directly. They can only be mentioned, if you like, not used; and so their existence will in general be presupposed in giving an account of a set.

KIT FINE

New York University

---

[23] It was only through adopting a platonist position that I was able to break away from the usual predicativist way of thinking and arrive at the present construction (I cannot help but echo Gödel's endorsement of the value of a platonist attitude as an aid to discovery!). The astute reader will notice a tension between the current theory of classes and the philosophy of procedural postulationism, as advocated in my paper, "Our Knowledge of Mathematical Objects," in John Hawthorne and Tamar Szabo Gendler, eds., *Studies in Epistemology*, Volume 1 (New York: Oxford, 2005), chapter 4. For how can classes be basic objects of mathematics if they are not *introduced* into the ontology, as procedural postulationism appears to require? It is not clear to me how the tension between these two views is best resolved.

A PUZZLE ABOUT DISBELIEF*

Although a variety of approaches to the semantics of belief ascription exist, something of a consensus has emerged in recent years. The consensus is that one of two favored approaches must be correct:

- An utterance exemplifying the surface form '*A* believes *p*' asserts a binary relation between a person *A* and a Russellian proposition *p*.[1]
- An utterance exemplifying the surface form '*A* believes *p*' asserts a ternary relation among a person *A*, a Russellian proposition *p*, and an implicitly-referred-to *mode of presentation*, or *guise*, of *p*.

In what follows I will present a puzzle for the former theory—the *Naïve Theory of Belief Reports*; I then go on to show how the puzzle applies with equal force to the latter theory—*Contextualism*.[2] As my title suggests, the puzzle involves (ostensible) reports of disbelief, or what I shall call, more neutrally, *negative* belief reports—sentences exemplifying the form '*A* doesn't believe *p*'. The puzzle, it should be noted, is gratifyingly simple; but its simplicity notwithstanding, it reveals a deep flaw in both proposals.

---

*This paper was delivered as the third Jerrold Katz Memorial Lecture at the City University of New York Graduate Center, on February 2, 2005. It was Jerry who, years ago, first got me interested in the semantics of direct reference and who, in his writings, in conversations, and in seminars, helped shape my understanding of the issues, even where we disagreed. (I think that he would be happy with at least some of what follows.) This paper is dedicated to his memory. I would like to thank the audience on that occasion for helpful comments and suggestions, and also to thank Jennifer Saul, Stephen Schiffer, an anonymous referee, and, especially, Ray Buchanan, for their valuable advice on an earlier draft.

[1] According to the Russellian, the proposition expressed by an atomic sentence '$\Phi\alpha_1, \ldots, \alpha_n$' can be represented by the ordered sequence $<<\alpha_1,\ldots,\alpha_n>, \Phi>$. $<<\alpha_1,\ldots,\alpha_n>, \Phi>$ is true just in case the sequence $<\alpha_1,\ldots,\alpha_n>$ is in the extension of the Φ relation. As can be gleaned from the representation, for each semantically-valued constituent, the proposition "contains" the semantical values of that constituent; in addition, it encodes the constituent's syntactical role (that is, as relation or as *i*-th *n*-tuple). The proposal extends to logically complex sentences, but the details need not detain us here.

[2] On the version of Contextualism to be discussed in the text—the Hidden-Indexical Theory of Belief Reports—context, rather than providing an individual mode of presentation, constrains what sorts of things can satisfy the mode-of-presentation parameter.

0022-362X/05/0211/573–93

I. THE PUZZLE

The *Naïve Theory of Belief Reports* (NT) consists in two independent claims:

(BIN) 'believes' expresses a binary relation between a person and a Russellian proposition.
(DR) Simple singular terms are directly referential.[3]

A consequence of NT is that obviously true (1) entails (2); but (2) contradicts obviously true (3):

(1) Lois believes that Superman flies.
(2) Lois believes that Kent flies.
(3) Lois does not believe that Kent flies.

Given that 'Kent' and 'Superman' co-refer, it follows that (1) and (2) must, according to (DR), relate Lois, in the same way, to one and the same proposition. (BIN) prohibits anything besides Lois and the ascribed proposition—a mode of presentation, say, or a sentence—to be relevant to determining the truth of (1) and (2). So (2) follows from (1). Since (3)—the negation of (2)—is true, the naïve theorist is faced with a contradiction.[4]

The naïve theorist's strategy to explain away the contradiction involves analyzing 'believes' in terms of a ternary relation, *BEL*, holding among a person, a proposition, and a mode of presentation or "guise" of that proposition. A useful gloss of '*BEL*(A, p, m)' is: A assents to p under guise m, where a guise is a sentence or sentence-like entity. It is in virtue of Lois's bearing *BEL* to the proposition that Superman flies—under some guise or other—that the ascription (1) is true. In effect, A's bearing *BEL* to p under some guise or other is a truthmaker for 'A believes p'.

The foregoing proposal, first described by Nathan Salmon,[5] is encapsulated in the following definition:

(EQ) Necessarily: A believes p iff [some m: A grasps p via m] (*BEL*(A, p, m))

According to (EQ), supplemented with our recent gloss, 'A believes p' is analyzable as the proposition that, for some guise m under which she grasps it, A assents to p under m.[6] Thus, for example, the fact reported by both (1) and (2) is just the fact that (1a) is the case (letting 's' refer to the proposition that Superman/Kent flies):

---

[3] Or, more to the point: the propositional contribution of a simple singular term is its referent.

[4] The standard formulation of Frege's Puzzle does not involve (3). It simply notes that (1) and (2) differ in truth value, even though they are equivalent modulo NT. As I show below, the strategy typically appealed to by the naïve theorist to resolve Frege's Puzzle cannot handle the current puzzle.

[5] *Frege's Puzzle* (Cambridge: MIT, 1986).

(1a) [some m: Lois grasps s via m] (BEL(Lois, s, m))[7]

I have suggested that the *BEL* relation should be viewed as similar to the *assent under a guise* relation. What does this amount to when guises are understood not as natural language sentences, but as internal representations—perhaps sentences in an internal code? Answering this question is important, because in certain situations we can speak truthfully of A's believing p even though there is no sentence A accepts whose content is p—perhaps there is not even a sentence he recognizes as having this content. I turn now to a brief discussion of the naïve theorist's answer.

I.1. Guises and Inward Assent. To see what the *assent under a guise* relation amounts to, let us assume that σ is a sentence in an internal code L.[8] In addition, let us help ourselves to the idea that σ "presents" p in L—that the proposition correlated with σ in L is p, or, formally, $L(\sigma) = p$.[9] The relation corresponding to inward assent (*assent under a guise*) is structurally similar to the familiar relation of outward assent (*assent under a sentence*) and can be modeled on it. The latter relation is mediated by acceptance of a sentence. That is, 'A outwardly assents to p' (or simply, 'A assents to p') is the product of two relations: (i) $E(S) = p$ (where $E(S)$ is a function from an English sentence S to the proposition expressed by S), and (ii) A accepts S:

---

[6] Since it refers to concepts an ordinary speaker cannot be assumed to possess, the naïve theorist, in endorsing (EQ), should not be viewed as thereby endorsing the claim that a speaker assertively uttering 'A believes p' thereby says (and recognizes herself as saying) what would be asserted by 'for some m under which A grasps p, A asserts to p as presented by m'. The official position of theorists such as Salmon is that the former "is analyzable as" the latter (see, for example, his "Being of Two Minds: Belief with Doubt," *Noûs*, XXIX, 1 (1995): 1–20, at p. 7), which falls short of claiming that the two forms express the same proposition (or, equivalently, that the former abbreviates the latter). To say anything stronger would blur the boundary between NT and Contextualism.

[7] Strictly speaking, Salmon's "grasping" condition in the quantifier matrix is redundant: it is a necessary condition for Lois's bearing *BEL* to s under m that she grasp s under m. As Salmon himself points out, the following principle (slightly rephrased) is partially definitive of *BEL* (he calls it "the *BEL* Principle"):

$BEL(A, p, m) \rightarrow A$ grasps p via m

(Salmon, "Being of Two Minds," p. 16, note 16.) While the grasping condition is redundant in (1a) it is essential if we are to be capable of expressing the possible situation at which A grasps p via m but nonetheless fails to bear *BEL* to p under m. And such situations—in particular, ones in which A withholds assent from p—figure centrally in what follows.

[8] It is not necessary that σ be a sentence—any representational state will suffice; but since nothing turns on the assumption, I will stick to it for the sake of familiarity.

[9] I suppress contextual parameters in the interest of readability.

(OA) A outwardly assents to $p =_{def} \exists S\ (E(S) = p\ \&\ (A \text{ accepts } S))$

(OA) seems to capture precisely the intuitive notion of (outward) assent. A definition of inward assent will parallel (OA). This relation is mediated not by A's acceptance of a natural language sentence, but by A's bearing an appropriate relation to an internal representation bearing the relevant content:

(IA) A inwardly assents to $p =_{def} \exists \sigma\ (L(\sigma) = p\ \&\ \mathbf{R}\ (A, \sigma))$

The relation $\mathbf{R}$ (in the second condition) holds between A and σ just in case σ is tokened as a belief, or, to borrow Stephen Schiffer's useful metaphor, just in case σ is "in" A's belief box. These are equivalent ways of saying that σ has, or has taken on, a certain causal/computational role—a role that it does not have simply in virtue of its syntactic properties.

So far, we have reduced inward assent to (something like) outward assent, replacing the acceptance condition in the former with the *tokened-as-a-belief* condition. I am not going to say more about what it means to be tokened as a belief, except to reiterate that it is typically unpacked in terms of σ's role in A's psychology (or A's "cognitive economy"). To say more would be to involve myself in unnecessary commitments and is in any case beyond the scope of this paper. But we have at least shown that the notion of inward assent to a guise is only as problematic as the notion of an internal state σ's having the content that p and that of σ's having the causal/computational role characteristic of belief. If these two notions present no insurmountable difficulties—which, it appears, they do not—then we can return to our topic.

II. A SOLUTION

How does analyzing a belief report as an existential generalization of the *BEL* relation allow the naive theorist to avoid the puzzle described in section I.1? Note first that the naïve theorist can take an utterance of (3) in one of three ways: as conveying disbelief in **s**, as conveying the negation of (2), or as conveying dissent from **s** (for convenience, the formal counterpart is displayed beneath each sentence):

(3a) Lois disbelieves **s**

[some m: Lois grasps $\neg$**s** via m] (*BEL*(Lois, $\neg$**s**, m))

(3b) There is no guise under which Lois assents to **s**

$\neg$[some m: Lois grasps **s** via m] (*BEL*(Lois, **s**, m))

(3c) There is some guise under which Lois dissents from s
[some *m*: Lois grasps s via *m*] ($\neg BEL$(Lois, s, *m*))

(3c) is synonymous with the more felicitous (3c′):

(3c′) Lois withholds assent from s.[10]

And (3c′), as (3c), leaves open the possibility that Lois assents to s under some other guise.

Among these options, the naive theorist can eliminate (3a) immediately. While (3) may in certain special circumstances give rise to the belief that Lois believes ¬s, it is not generally the case that the "doesn't believe *p*" idiom conveys belief in $\neg p$. (For example, the case in which *A* doubts *p*—that is, in which *A* believes neither *p* nor its negation—is one in which a report of the form '*A* doesn't believe *p*' is obviously true, but such a case is clearly not one in which *A* can be said to believe $\neg p$.) Thus we need only consider (3b) and (3c).

The naïve theorist claims that there is a (nonsemantic) regularity according to which speakers use '*A* doesn't believe *p*' to communicate that *A* withholds assent from *p*. The suggestion is that a speaker, in uttering (3), will typically intend to convey, not (3b), but (3c). Since (3c) is compatible with (1a), this proposal offers an explanation of our intuition that a speaker who utters both (1) and (3) does not thereby contradict himself. As Jennifer Saul writes:

> What is it we are trying to convey when we utter (3)? We are trying to convey that there is some guise, which Lois understands, under which she *does not* accept the proposition in question. We are trying, then, to

---

[10] Salmon takes (3c′) to be equivalent to the claim that Lois withholds belief from s (see Salmon, *Frege's Puzzle*, p. 111; "Being of Two Minds," p. 7). But, on the face of it, this is incompatible with (BIN). While speaking of 'withholding *assent* from s' is unobjectionable, since 'assent' (at least as we have been using it) expresses a ternary relation, speaking of 'withholding belief' *is* objectionable, since, so far as the naïve theorist is concerned, 'believes' expresses a binary relation. Strictly speaking, '*A* believes *p* under *m*' should not be acceptable. (Unless the role of 'under *m*' is not to indicate an argument position in the *believes* relation but to modify how the relation obtains—that is, unless 'under *m*' functions as an adjunct. But it is doubtful that the naïve theorist would find such an analysis congenial, since the best way of implementing such a strategy—involving quantification over events—would require dispensing with (BIN).) If we are to comply with (BIN) we can only speak of *believing* [$R(A, p)$], *not believing* [$\neg R(A, p)$], *disbelieving* [$R(A, \neg p)$], *not disbelieving* [$\neg R(A, \neg p)$], and truth-functional combinations thereof. Within these limitations, there is no possibility of defining 'withholding belief'. Of course the naïve theorist is free to *use* the 'withholds belief' idiom as interchangeable with 'withholds assent'. But to the extent it sounds natural to speak in this manner, it is because *believing*, like *assenting*—like *BEL*, in fact—is ternary. A naïve theorist such as Salmon should be the last person to endorse use of this idiom.

Having said this, I will continue to use the 'withholds belief' and 'withholds assent' interchangeably, since I am not troubled by the implication mentioned above.

> convey that she withholds belief from the proposition. If there is such a guise, it is true that Lois withholds belief from the proposition that Superman can fly. Salmon and Soames maintain that we mistake (3), which must be false if (1) is true, for (3c)—the implicated proposition—which may still be true.[11]

The claim here is that speakers mistake the proposition implicated by an utterance of (3)—namely, (3c)—for the proposition it literally expresses—namely, (3b).

It is worth pausing for a moment to admire this strategy, as it contains an implicit diagnosis as to why a speaker makes the mistake in question. The diagnosis is that the speaker mistakes an internal negation for an external one. This explains, at least in part,[12] how a speaker misidentifies the contents of a negative belief report and, consequently, why it is that he does not take the reports (1) and (3) to contradict one another. Again, the speaker has misunderstood where the negation occurs relative to the quantifier, inadvertently committing a simple scope error—a common occurrence.

The NT advocate can thus be seen as endorsing the following thesis:

> Although '*A* doesn't believe *p*' expresses the proposition that *A* fails to assent to *p* under any guise, speakers typically use it to communicate the proposition that *A* withholds assent from *p*.[13]

Attractive as it is, however, there are serious problems with this strategy. I turn to these now.[14]

III. PRAGMATICS AND WISHFUL THINKING

The first problem for this approach is obvious. If the naïve theorist is to avoid arbitrariness, it must be impossible, on her view, to communicate the proposition that there is no guise under which *A* assents to *p* by uttering a sentence of the form '*A* doesn't believe *p*' (or variants thereof). Unless a sentence of the form '*A* doesn't believe *p*' is always used to convey that *A* withholds belief from *p*, we must rely on a "wishful thinking strategy"—in effect: interpret '*A* doesn't believe *p*' as '*A* withholds belief from *p*' exactly when doing so avoids embarrassing counterexamples to NT.[15] Obviously this is not a serious proposal, so the naïve theorist must bite the bullet and accept the consequence just described—that negative belief reports are always used to convey withholding of belief. Call this the "Bite-the-Bullet" consequence:

---

[11] "The Pragmatics of Attitude Ascription," *Philosophical Studies*, XCII, 3 (1998): 363–89, at p. 372. (I have changed the numbering.)

[12] The proposal explains how the error arises; it fails to explain the prevalence of the error. This will be explained presently.

[13] For *S* to express *p* at a context **c** is just for *p* to be the value, at **c**, of the character function assigned to *S*. Equivalently: what *S* expresses at **c** is what a speaker can use *S* to *say* at **c**.

[14] An alternative strategy would claim that a use of (3) will inevitably involve an implicit restriction to a domain D of mild-mannered-reporter-involving guises, as in ($3b_D$):

($3b_D$) [no *m*: Lois grasps s via $m]_D$ (*BEL*(*Lois*, s, *m*))

Since ($3b_D$) is compatible with (1a), contradiction is avoided. The naïve theorist who finds this congenial has unwittingly become a Contextualist, as this strategy is merely a notational variant of the Contextualist proposal to be considered below (section VIII)—namely, the Hidden–Indexical Theory.

(BB) Always interpret '*A* doesn't believe *p*' as '*A* withholds assent from *p*'.

It seems quite clear, however, that a speaker can use '*A* doesn't believe *p*' to deny categorically that there is *any* guise under which *A* believes *p*. Otherwise there is no explanation of the fact that, in the following exchange, Jones's utterance contradicts Smith's, nor of the fact that Jones asserts a falsehood:

Smith: Tony Blair believes that Mick Jagger was knighted.
Jones: (*Incredulously*) Hogwash! Tony Blair doesn't believe that Mick Jagger was knighted.

On the current account, Jones's utterance does not contradict Smith's, since the propositions it assigns to their respective utterances are consistent. This can be seen by rendering the exchange in accordance with (BB):

Smith: There is some guise *m* of *the proposition that Mick Jagger was knighted* such that Tony Blair assents to *it* under *m*.
Jones: There is some guise *m* of *the proposition that Mick Jagger was knighted* such that Tony Blair does *not* assent to *it* under *m*.

In addition, Jones does not assert a falsehood on the current account, since the proposition it has him asserting is true. After all, there surely is some guise of the proposition that Mick Jagger was knighted under which Blair does not recognize it. Evidence for a guise presenting Jagger in a manner unfamiliar to Blair is sufficient to show this. An example might be a guise which presents Jagger as impersonating Karl Lagerfeld on *Saturday Night Live*. Blair would not assent to the proposition thus presented.[16]

---

[15] On the idea of wishful thinking, I am indebted to the excellent but otherwise irrelevant paper by Cian Dorr, "Non-Cognitivism and Wishful Thinking," *Noûs*, XXXVI, 1 (2002): 97–103.

Suppose, however, that there exists an apparently irrelevant guise of the proposition that Mick Jagger was knighted under which Blair happens to believe it (perhaps phone wires get crossed, and he hears Jagger boast of his recent knighthood).[17] On the view being considered, Blair's believing this proposition under such a guise would contradict Jones's utterance. Yet it might be objected that Jones would not take such a circumstance to contradict his assertion—after all, he did not have such a guise in mind—and that consequently (BB) gets things right after all. But there is reason to think that Jones would, when faced with this state of affairs, retract his statement. Indeed, if the irrelevant guise is the sole basis for Smith's assertion, then, although he spoke misleadingly (implicating that Blair would accept the content sentence), he did not speak falsely. In fact, were Smith to present the guise in question as evidence for his assertion, it seems clear that Jones would have to accept Smith's claim as true. (Again, it is not being denied that Smith's *utterance* was misleading, but the naïve theorist can easily explain this by appealing to the above-cited implicature.)

The following case, borrowed from Mark Crimmins,[18] provides further evidence against the current theory's plausibility: Gonzales tells me that there is some guise under which I know him *and* under which I believe that he is an idiot. Since (i) I believe that Gonzales is clever (and thus *not* an idiot) under his current guise, and (ii) since I am incapable of determining which guise in particular is the idiot-presenting one, and finally (iii) since I trust him under his current guise, it seems that I can truthfully say 'I falsely believe that you are an idiot'. If, on the other hand, I know that Gonzales is mistaken (perhaps, being unreasonably charitable in interpreting others' behavior, I do not believe *anyone* to be an idiot) I can respond to him by uttering the following: 'You're wrong—I don't believe you are an idiot.' In uttering this I am not intending to convey *I withhold belief as to whether you are an idiot,* since this would not express disagreement. Rather, I am intending the following: *There is no guise under which I believe that you are an idiot.* But this would not be possible under the strict conventions surrounding negative belief reports apparently forced on us by the current proposal.

---

[16] Since grasping p under guise m is a necessary condition for assenting to p under m. Thus, Blair's failure to grasp the relevant proposition under the Lagerfeld guise precludes his assenting to it under that guise. Note that the argument from unfamiliarity to nonassent assumes Salmon's *BEL* Principle, described in note 7.

[17] I am indebted here to an anonymous referee.

[18] "I Falsely Believe That *P*," *Analysis*, LII (1992): 191.

Thus, (BB) must be rejected. Should we reconsider the wishful thinking strategy? Obviously not. In fact, the naïve theorist has an alternative to the latter that provides a principled way of explaining how our speaker manages to convey the relevant proposition when uttering (3). She offers the following proposal:

(P) A belief reporter communicates (3c) in uttering (3) only when it is mutually known that Lois believes **s**—that is, only when it is mutually known that Lois assents to **s** under some guise or other; otherwise, in uttering (3) the belief reporter communicates (3b).

This suggestion would appear to result from a simple application of Grice's Maxim of Quantity, "Make your contribution as informative as is required (for the purposes of the exchange),"[19] or, alternatively, of Stephen Levinson's Q-principle: "Do not provide a statement that is informationally weaker than your knowledge of the world allows."[20] If there is no good reason to convey the (comparatively) weak claim that Lois withholds assent from **s**, then one should convey the stronger; otherwise, the weaker claim prevails.

IV. IMPLICATURE VERSUS UNDERDETERMINATION

Given that (P) describes a regularity, it makes sense to view the interpretation of (3) in terms of (3b) as a generalized conversational implicature. According to (P), the default interpretation of (3) is (3b), where this interpretation is cancelled only in contexts in which it is clear that Lois assents to **s** under some or other guise. Whereas a *particularized* conversational implicature occurs only when a literal construal of my utterance constitutes a prima facie violation of a given conversational maxim, a generalized conversational implicature does not require such a violation: such implicatures are, in the normal course of things, carried by certain sentences, requiring no specific cues in order to get generated.[21] As such, a generalized conversational implicature may be regarded as a "default" or "preferred" interpretation, similar to conventional linguistic meaning in requiring no contextual input, dissimilar in being cancellable.[22] In each case, the existence of the implicature depends upon general adherence to conversational maxims. In the generalized case, the existence of the implicature is a direct consequence of the assumption that certain of the maxims are being observed (together with facts about the literal meaning of the relevant utterance), and requires no contextual input; such input is required only to cancel the implicature. In the particularized case, existence of the implicature depends, crucially, on contextual input—specifically, on the fact that a maxim is explicitly being flouted.[23]

---

[19] Grice, "Logic and Conversation," in his *Studies in the Way of Words* (Cambridge: Harvard, 1989), pp. 22–40, at p. 26. As is well known, compliance with this maxim comes under the general requirement imposed by the "Cooperative Principle": "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged."

[20] Levinson, *Presumptive Meanings: The Theory of Generalized Conversational Implicature* (Cambridge: MIT, 2000), p. 76. Although I appeal to Grice's maxim in what follows, this is merely for the sake of familiarity; appeals to the Q-principle would have been equally effective.

[21] See Grice, "Logic and Conversation," pp. 37–38. Levinson takes generalized conversational implicatures to constitute a level of meaning (what he labels "utterance type meaning") distinct from literal meaning on the one hand and utterance meaning on the other. They are default meanings, so to speak, attaching to sentence types in virtue of certain pragmatic principles but also readily cancellable. These principles, or "heuristics," are similar in content to Grice's maxims of manner and quantity (pp. 35–42).

Yet the characterization of (P) as a generalized conversational implicature is after all not quite apt. While (3b) is a default interpretation of (3), (3b) cannot really be said to be an implicature, since there is no proposition assigned to (3) that actually implicates (3b). According to (P), (3b) is simply what one means when uttering (3) in those contexts where it is not mutually known that Lois believes **s**. Because of this the shift that takes place when it is mutually known that Lois believes **s** is not happily described as being effected by cancellation of an implicature—there is no proposition we can identify whose implicature has been cancelled.

These considerations suggest an alternative reading of (P). On this reading, the linguistic meaning associated with the sentence-type (3) has no determinate logical form and consequently no determinate propositional content. The assignment of logical form to an utterance of (3) is determined by contextual factors—in actuality, a single contextual factor: whether the proposition that Lois assents to **s** under some guise or other is mutually known. In a context at which this proposition is mutually known an utterance of (3) expresses (3c); otherwise it will express (3b).

To use familiar terminology, the proposal is that the linguistic meaning of (3) *underdetermines* logical form and thus what is said;

---

[22] The terminology is Levinson's. His discussion of the distinction is useful; see Levinson, pp. 16–17.

[23] For example, my uttering 'It's getting late' to a friend at a party to convey the proposition that I want to go home is a particularized conversational implicature; the sentence does not typically convey the information that the speaker wishes to go home. Since the time itself is presumably of no particular interest in the circumstance, determining what I intend involves, among other things, figuring out why, although apparently cooperative, I am offering irrelevant information. On the other hand, while my use of the sentence 'Jones bought a gun and robbed a bank' also generates an implicature (namely, that Jones bought a gun and that he subsequently robbed a bank), recovering the intended meaning does not call for any special calculation on the part of the hearer. It is thus a generalized conversational implicature.

logical form can be assigned only if we are given relevant contextual input.[24] This is not an ad hoc maneuver: it is often claimed that the linguistic meaning of a sentence such as 'I read John's book' underdetermines what is said by an utterance of that sentence. While the linguistic meaning of this sentence is compatible with a number of propositions (for example, *I read the book John wrote, I read the book John owns, I read the book John reviewed,* and so on), a speaker will have said at most one such proposition in uttering the sentence. Without contextual input to flesh out the linguistic meaning—in particular, to specify which sense of 'John's book' is relevant—no proposition will be determined. The approach can be seen to be fruitfully applied to a variety of sentence types. I am suggesting that the naïve theorist apply it to negative belief reports (this would require abandoning the suggestion that (3c) is somehow implicated by certain utterances of (3)).

Although nothing turns on which interpretation of (P) is selected, I will assume in what follows the interpretation according to which (3)'s linguistic meaning underdetermines its logical form.

V. INDEPENDENT MOTIVATION FOR (P)

Before determining whether (P) can resolve the puzzle presented in section I, it should be noted that there is reason for a naïve theorist to hold (P), however it is interpreted—that is, reason independent of the puzzle presented in section I. (P) is a consequence of a well-known generalized conversational implicature, one that is justified by the Maxim of Quantity. The implicature I have in mind obtains between my assertion that *some Fs are G* and the proposition that, for all I know, *some Fs are not G.* For example, my asserting (4b) encourages the belief that, for all I know, (4a) is true. At a stage of the conversation where each of (4b)–(4d) are relevant, my contributing (4b) will be expected to give rise to the belief that I have not ruled out the possibility of (4a)'s being true. Had I ruled out (4a) as a possibility, I would have uttered (4d) instead.

(4a) Someone didn't pass the final exam.
(4b) Someone passed the final exam.
(4c) No one passed the final exam.
(4d) Everyone passed the final exam.

To see this, let us assume that we have arrived at a stage in our conversation where (4b)–(4d) become relevant. (Perhaps you have asked 'Did any students pass the final?') In addition, let us also assume that I have only graded a few exams, and that all of these are "passes." The Maxim of Relation ("be relevant") demands that I choose from (4b) – (4d)—otherwise I am changing the subject.[25] The Maxim of Quality demands that I choose (4b)—otherwise I am trafficking in falsehoods or claims for which I lack adequate evidence. Since I am expected to be sufficiently informative—since compliance with the Maxim of Quantity is assumed—I have encouraged the (true) belief that I am not in a position to reject (4a).[26]

---

[24] See François Recanati, *Literal Meaning* (New York: Cambridge, 2004), pp. 56–57. The view derives from Dan Sperber and Deirdre Wilson, *Relevance: Communication and Cognition* (Cambridge: Blackwell, 1995, second edition).

Given that compliance with Relation and Quality are generally taken for granted, what directly underwrites the implicature (in the circumstance) is the assumption that I am complying with the Maxim of Quantity—in effect, with the expectation that I am being maximally informative.[27]

The "implicature rules" apparently at work here can be represented as follows:

(5a) $[\text{some } x\text{: } Fx]\ (Gx) \gg \Psi\ [\text{some } x\text{: } Fx]\ (\neg Gx)$
(5b) $[\text{some } x\text{: } Fx]\ (\neg Gx) \gg \Psi\ [\text{some } x\text{: } Fx]\ (Gx)$

(letting 'Ψ' stand for the operation 'for all the speaker knows...' and '$\gg$' stand for 'implicates...unless there is good reason to the contrary'). Note that if the implicature fails, for any reason, this is because the context contains the proposition that the implicated proposition is false.[28] For example, (5a) will fail if it is part of the context that the speaker knows that $[\text{all } x\text{: } Fx]\ (Gx)$; (5b) will fail if it is part of the context that the speaker knows that $[\text{all } x\text{: } Fx]\ (\neg Gx)$.[29]

---

[25] Grice, "Logic and Conversation," p. 27. To make this more plausible, assume that you do not know the identities of the students, so that sentences such as 'Jones passed' would not be relevant.

[26] I am for the moment ignoring claims involving statements with cardinality quantifiers, for example:

(4e) At least five students passed the exam.

These are relevant, if, perhaps overly informative; but including them does not affect my point. A parallel argument covers such cases. For example, say I have graded five exams, and all the exams are passes. I answer your question with (4e) instead of (4b). The same expectation guarantees—all things being equal—that I am not in a position to reject (4a).

[27] Parallel reasoning reveals that my asserting that some *F*s are *not G* implicates that, for all I know, some *Fs are G*.

[28] For a context c to *contain* a proposition p is for p to be mutually known among the conversational participants at c.

[29] (5a) and (5b) are thus examples of scalar implicatures. Such implicatures are generated at a context c at which each member of a set of propositions $\{p_1, p_2,..., p_n\}$ is relevant, and where the members are arranged so that each p_{i+1} intuitively "strengthens" p_i. Asserting a proposition p_k weaker than p_n at c generates the implicature that, for all the speaker knows, $\neg p_n$ (indeed, it implicates each member of

The corresponding "rule" governing negative belief reports is simply an instance of (5b):

(6) [some m: A grasps p via m] ($\neg BEL(A, p, m)$) » Ψ [some m: A grasps p via m] ($BEL(A, p, m)$)

That is, my assertion that *A* withholds assent from *p* implicates that, for all I know, *A* (also) assents to *p*. As before, if the implicature fails, this is because the context contains its negation—that there is no *m* under which *A* both grasps, and accepts, *p*.

(6) provides the naïve theorist with a powerful independent basis for (P). If we suppose, with the naïve theorist, that belief reports are to be analyzed in terms of an existential quantification, then we should expect that negative belief reports will, in certain contexts, give rise to the implicature described in (6).[30] As (6) makes clear, asserting (3c)—that Lois withholds assent from **s**—will implicate that, for all the speaker knows, Lois assents to **s**. But there may be circumstances in which, although (3b) is not mutually known, the speaker believes (3b). To assert (3c) in such a circumstance will license the belief that, for all the speaker knows, Lois assents to **s**. Since he wants to avoid giving rise to false beliefs (that for all he knows, Lois does assent to **s**), he communicates (3b). This allows him to say something true without endorsing a falsehood.

VI. A PROBLEM

(P) thus allows us to steer a course between the "wishful thinking strategy" on the one hand and (BB) on the other. If a negative belief report *S* typically expresses withholding of assent, then (P) is a principled way to avoid choosing between the extreme alternative that *S always* expresses belief-withholding and the equally undesirable alternative that *S* weakens arbitrarily. This is because there is

---

$\{\Psi \neg p_i \mid k < i \leq n\}$). See the appendix to Scott Soames, "How Presuppositions Are Inherited: A Solution to the Projection Problem," in Steven Davis, ed., *Pragmatics: A Reader* (New York: Oxford, 1991), pp. 428–70 (esp. pp. 446–50, 454–57) for a useful discussion. (An important feature of scalar implicatures is that the operative expressions (the items in the scale) are lexically simple (see Levinson, *Presumptive Meanings*, pp. 79–81). For example, in (5a), the expressions are {*some*, *all*}, both lexical determiners. A speaker who uses *some* indicates that he is not in a position to use *all* (unless the context indicates otherwise); she thus implicates that for all she knows use of *all* would yield a falsehood. Although the lexicalization constraint is arguably met in the scale relevant to (6) below (consider {*dissents*, *disbelieves*}), the NT advocate can adopt (6) whether or not it is a genuine scalar implicature. It is enough that it is variously underwritten by the Maxim of Quantity and the Q-principle.)

[30] Indeed, if she were to accept (BB), the naïve theorist should expect this implicature to accompany regularly negative belief reports.

independent reason to think (P) will obtain if, as the NT-theorist claims, '*A* doesn't believe *p*' is regularly used to communicate the proposition that, for some *m* via which *A* grasps *p*, $\neg BEL(A, p, m)$. Given the implicatures described in (5a/b), and assuming the naïve analysis of belief reports, we should expect that negative belief reports will give rise to the implicature described in (6).

Thus, the NT-advocate has a principled response to the potential charges of extremism on the one hand and arbitrariness on the other. But things are not well. To see this, consider a context in which we *reverse* the order in which the ascriptions (1) and (2) are made. For example: Lois meets Kent before she learns of Superman's existence (or, if you prefer, learns of Kent's alter ego).[31] In uttering (3) at this point in time, our speaker would assert (3b). According to the naïve theorist's principle (P), and, more fundamentally, according to the recently described "rule" (6), asserting (3c) at *t* would implicate the proposition that, for some *m* via which she grasps **s**, *BEL*(Lois, **s**, *m*). Since our speaker does not believe the latter proposition—indeed, he believes it to be false—he would not intend anything that would license his audience's belief in it.

To continue our story: Lois subsequently meets Superman and learns of his ability to fly. Our belief reporter, in describing Lois's cognitive state, now utters (1)—which, of course, is content-identical to (2). Yet in so doing our belief reporter has contradicted the proposition conveyed by his former report—namely, (3b). Unfortunately, we cannot go ahead and revise his intention in uttering (3) after the fact. And while it can be argued that any *future* utterance of (3) will be construed as a report of belief withholding—namely, as (3c)—this does nothing to mitigate the contradiction currently staring us in the face.[32]

It will be noted that, strictly speaking, there is no *formal* contradiction here: (3) asserts that, relative to a given time *t*, Lois fails to believe **s** under *any* guise; (2) asserts that, relative to subsequent *t′*, there is some guise under which she believes it. These two propositions are fully compatible. Yet, it is no part of the story just related that, in the interval between *t* and *t′*, Lois acquires any information that would make it plausible that she comes to accept 'Kent flies'. So it would be arbitrary for the NT-advocate to assume that Lois somehow changes her Kent-relevant beliefs in the interval (such that, for example, the utterance of (2) would be acceptable). And it goes without saying that it would be question-begging for him to assume, due to its equivalence with (2), that the truth of (1) at *t′* sufficed for such a change in Lois's beliefs.[33]

---

[31] For all I know, this is how things (fictionally) actually happened.

[32] In an excellent recent discussion, Sean Crawford argues that the NT-advocate can express belief-withholding with the resources of the binary *believes* relation and that thus "there is no need for the Russellian to 'go ternary' and invoke 'proposition guises'" ("A

VII. THE PRAGMATIC STRATEGY REVISITED

As indicated above (see note 4), a puzzle can be formulated for NT using only (1) and (2). According to NT, (1) and (2) express the same proposition. Yet, intuitively, they differ in truth value: (1) is true, whereas (2) is false. The resolution is well known: instead of dropping or modifying (DR) or (BIN), the naïve theorist denies that (1) and (2) differ in truth value. What the naïve theorist claims is that, in uttering (1) a speaker conveys (1*f*), whereas in uttering (2) a speaker conveys

---

Solution for Russellians to a Puzzle about Belief," *Analysis*, LXIV, 3 (2004): 223–29, on p. 227). On Crawford's analysis, for Lois to withhold belief in s is simply for Lois to believe that she does not believe s. This situation is rendered semi-formally as:

(3c^{B}) Believes(Lois<<Lois, s>, does not believe>).

While (3c^{B}) may be an effective replacement for (3c), it is equally powerless to resolve the contradiction currently faced by the NT-advocate. If, in uttering (3) at *t* our belief reporter intends (3c^{B}), as opposed to ¬Believes(Lois, s), he will generate the (false) implicature that the proposition, Believes(Lois, s), is true. That this is so should already be obvious given Crawford's paraphrase of (3c^{B}): to assert that Lois believes that she does not believe that Kent flies is to license the inference that, for all the speaker knows, Lois does in fact believe it. This is made clear by the following exchange:

A : Just out of curiosity, does Lois believe that Kent flies?
B : Lois thinks she doesn't believe that Kent flies.

B's utterance clearly gives rise to the belief that Lois believes that Kent flies. Additional evidence for this claim comes from the following:

(a) The implicature can be cancelled: *Bs* response can be completed by '...but of course she does' without contradiction;
(b) Asserting the implicated proposition is not redundant: completing *Bs* response by '...and in fact she doesn't' provides new information.

Note that, while the implicature is not scalar, as it would be were we to assume (3c), it can be seen to require a similar appeal to Grice's Maxim of Quantity.

(One might ask, How could it be the case that *S* can believe *p* and yet also believe that she does not believe *p*? The naïve theorist we have been considering has an answer: under certain ways of grasping it, *S* does not recognize the belief she has. This possibility clearly falls out of the ternary analysis. On the other hand, it must remain a primitive feature of the binary analysis that one can simultaneously believe, and yet not take himself to believe, one and the same proposition. Thus, although the ternary analysis is more complex, it has a distinct explanatory advantage over the binary analysis.)

[33] I am indebted to Schiffer for discussion helping to clarify the points made in this paragraph.

(2*f*); he adds that, when reporting a belief, a speaker typically confuses what his utterance implicates with what his utterance says:

(1*f*) *BEL*(Lois, **s**, *f*('Superman flies'))
(2*f*) *BEL*(Lois, **s**, *f*('Kent flies'))

Recall that a guise is a sentence-like entity through which a speaker grasps a proposition. What (1*f*) states is that Lois believes **s** under a guise that is similar to 'Superman flies' (where *f* is a function from a sentence to an appropriately similar guise); what (2*f*) states is that Lois believes **s** under a guise that is similar to 'Kent flies'.[34] The naïve theorist maintains that, in uttering (1), I implicate (1*f*) and, in uttering (2), I implicate (2*f*). Thus, the intuition that (1) and (2) differ in their truth-conditional content derives from mistakenly identifying what utterances of these sentences implicate with what they say. While what they say is, strictly speaking, identical, what they implicate differs, and this, we are told, accounts for our intuitions.

I do not wish to evaluate this strategy here;[35] I introduce it merely to consider its application to the current problem. Perhaps our speaker, in uttering (3), attempts to convey not (3b) but (3*f*):

(3*f*) $\neg$*BEL*(Lois, **s**, *f*('Superman flies'))

That is, perhaps he attempts to convey the proposition that Lois does not believe **s** under a guise similar to 'Superman flies'. This prop-

---

One strategy for avoiding the contradiction is to appeal to the Maxim of Relation. On this view, to assume that our speaker in uttering (3) at *t* intends (3b) would have him violate this maxim by irrelevantly quantifying over all guises, as opposed to restricting the quantifier to all relevant (that is, mild-mannered-reporter-involving) guises, as in (3b$_{\mathrm{D}}$):

(3b$_{\mathrm{D}}$) [no *m*: Lois grasps **s** via *m*]$_{\mathrm{D}}$ (*BEL*(Lois, **s**, *m*))

(where D consists exclusively in mild-mannered-reporter-involving guises). In a context in which it is mutually known that Lois assents to **s** under a superhero-involving guise, a speaker, in uttering (3), may implicitly restrict the quantifier *à la* (3b_{D}) and thus avoid contradicting himself. As I mentioned in note 14, a naïve theorist advocating this approach has unwittingly become a Contextualist. In addition, it is unclear why only mild-mannered-reporter-involving guises would be relevant in the context described in the text. It seems that such guises would be exclusively relevant only in a context in which there is some probability that Lois assents to **s** under an alternative guise. But the imagined context contains no evidence for this claim.

[34] The function should be understood as relativized to a speaker and time: the guise Lois associates with 'Superman flies' may differ from the guise that Jimmy Olsen associates with it, and may differ from the guise she herself associated with it before she learned the Superman could fly.

[35] But see Mitchell Green, "Direct Reference and Implicature," *Philosophical Studies*, xci, 1 (1998): 61–90, and Saul, "The Pragmatics of Attitude Ascription," for critical discussion.

osition is fully compatible with (2*f*). So, under the suggested analysis, our speaker has not contradicted herself after all.

Once again, the main difficulty with this proposal is that it would have our speaker violate the Maxim of Quantity. If the speaker knows that Lois fails to believe **s** under any guise—that is, if he knows (3b)—then he should convey this, and not the weaker (3*f*), which claims only that Lois fails to believe the proposition under a particular guise—that of *f*('Superman flies').

Let us take a closer look at this response. (3*f*) states that Lois assents to **s** under *f*('Superman flies'). Since our speaker believes that, for all *m* under which Lois grasps **s**, Lois withholds assent from **s** under *m*, his asserting (3*f*) would contribute a proposition weaker than context requires.[36] Such a contribution would constitute a violation of the Maxim of Quantity. In asserting this relatively weak claim he presents himself as leaving open the possibility that, for some *m* under which Lois grasps **s**, *BEL*(Lois, **s**, *m*). But this is false: according to our scenario, he is not willing to entertain this as a possibility.

Given that the speaker has done nothing to prevent the audience from making the obvious inference, his utterance licenses the belief that he is willing to entertain the possibility that there is some *m* such that Lois grasps **s** via *m* and such that *BEL*(Lois, **s**, *m*). Our assumption throughout has been that our speaker is well informed and well intentioned—in addition to being linguistically competent, he knows the relevant facts and has no incentive to mislead. But this assumption, combined with our recently-described scenario and with the proposal currently on the table (and assuming familiar pragmatic principles), leads to a contradiction.

The NT advocate thus faces a dilemma: either accept the wildly implausible claim that it is impossible to use 'Tony Blair doesn't believe that Mick Jagger was knighted' to contradict 'Tony Blair believes that Mick Jagger was knighted', or face the unwelcome consequence that our well-informed and well-intentioned belief reporter has no option but to contradict himself in reporting Lois's beliefs in an unexceptionable situation (assuming, of course, that he is conforming to Grice's Cooperative Principle).[37]

VIII. CONTEXTUALISM TO THE RESCUE?

NT's main competition, Contextualism, was designed to provide an alternative to NT's account of Frege's Puzzle. Perhaps, on this approach, the difficulty presented by our recent example disappears. Although I will focus on a specific Contextualist approach, namely the *Hidden-Indexical Theory* (HIT) as it is developed by Stephen Schiffer,[38] my discussion applies with equal force to the closely related proposal of Mark Crimmins and John Perry.[39]

---

[36] What our speaker believes, (3b), is equivalent to [all *m*: Lois grasps **s** via *m*] (¬*BEL*(Lois, **s**, *m*))—that is, [no *m*: Lois grasps **s** via *m*] (*BEL*(Lois, **s**, *m*)).

[37] Of course, the contradiction is not literal, since I am assuming that the proposed convention is nonsemantic, but that does not make it any less of a contradiction.

HIT dispenses with both (BIN) and (EQ). Instead of claiming with NT that a sentence such as (1) is necessarily equivalent to an existential quantification of *BEL*(Lois, **s**, *m*), HIT claims that the very content of an utterance *u* of (1)—the proposition that *u* expresses—is the existential quantification of a ternary relation very much like *BEL*. More precisely, what HIT claims is that 'believes' stands for a ternary relation among a person, a proposition and a mode of presentation—a relation which we will refer to as the "*believes* relation," or, simply, "*believes*." In fact, for present purposes, we can take *believes* and *BEL* to be one and the same relation, bearing in mind that, under HIT (but not NT), *believes* is the relation expressed by the English word 'believes'.

According to HIT, an *utterance* of (1) asserts the following:

(1^H) [some m: $\Phi^* m$] (Believes(Lois, **s**, m))

(where Φ^* is a contextually-determined property of modes of presentation).[40] Some remarks about (1^H) are in order. First, it locates in belief reports an implicitly referred-to condition on modes of presentation; the content of an utterance of (1) is thus not fully articulated at surface syntax. In addition, different utterances of the same report will vary in their implicit references, constraining the mode-of-presentation parameter in different ways.[41]

Intuitively, (1^H) states that there is a guise meeting contextually-determined condition such-and-such under which Lois believes **s**. The imagined utterances (1) and (2) differ in point of their implicit references to relevant mode-of-presentation properties—what I shall refer to as their *styles*. The style implicitly referred to in the utterance of (1) presents **s** as superhero-involving, whereas the style implicitly referred-to in the utterance of (2) presents **s** as mild-mannered-reporter-involving. This accounts for the intuition ordinary speakers have that (1) and (2) differ in terms of their truth conditions. More importantly for present purposes, we can account for the intuition ordinary speakers have that the proposition expressed by the negation of (2) does not contradict the proposition expressed by (1)—(3^H) and (1^H), respectively, according to HIT.

---

[38] "Belief Ascription," this JOURNAL, LXXXIX, 10 (October 1992): 499–521.

[39] "The Prince and the Phone Booth: Reporting Puzzling Beliefs," this JOURNAL, LXXXVI, 12 (December 1989): 685–711; Crimmins, *Talk about Beliefs* (Cambridge: MIT, 1992).

[40] I will use Φ^* an *explicit* indexical, one whose semantical value is a contextually definite mode-of-presentation property. Thus, according to my usage, (1) and (1^H) do not differ in point of context-sensitivity; they differ only insofar as (1^H) makes explicit the respect in which (1) is context-sensitive. It is not itself context-*in*sensitive.

[41] Another interesting feature—that the argument place occupied by modes of presentation is not realized at surface structure—need not detain us here.

(3^H) $\neg$[some m: $\Phi^* m$] (Believes(Lois, **s**, m))[42]

Assuming our recent gloss on (2), what (3^H) states is that there is no mild-mannered-reporter-presenting style under which Lois believes **s**. While this proposition contradicts the suggested analysis of (2) it is fully compatible with (1). Thus, (1)–(3) are assigned their intuitively correct truth conditions and the puzzle they presented to NT (in section I.1) is avoided.

While style-reference can be invoked in order to block problematic substitutions, the HIT advocate also maintains that, in certain cases, style reference effectively drops out. In such cases, NT and HIT will agree on the truth conditions they ascribe to the relevant utterance. Indeed, a stated advantage of HIT is its flexibility: it can constrain the mode of presentation parameter as coarsely or as finely as our truth-conditional intuitions demand. While it can constrain the parameter sufficiently finely to resolve the puzzle discussed in the text, it can also eliminate the constraint when necessary. For example, an utterance of Schiffer's sentence (7), intuitively, provides no substantive constraint:

(7) Everyone who has ever met her has believed that Madonna has musical talent.

Since the proposition that Madonna has musical talent is not presented to everyone who has ever met her in the same way—and since speaker and hearer are likely aware of this fact—the style will be trivially satisfiable; for example, we can suppose it to impose a condition of self-identity:

(7^H) [every y: y has met Madonna]
([some m: $\Phi^* m$]
(Believes(y, the proposition that Madonna has musical talent, m)))

Thus (7^H), suitably interpreted, will deliver the intuitively correct truth conditions for (7).

---

[42] The notational variant of HIT mentioned in note 14 would restrict the quantifier domain to the set of modes of presentation satisfying the contextually salient style.

Acceptance of HIT requires an adjustment to (P):

> (P^H) Although a belief reporter always communicates (3^H) in uttering (3), the contextually-determined style, Φ*, is nontrivially satisfiable only when it is common knowledge that Lois believes **s** under some mode of presentation or other—that is, only when it is common knowledge that some instance of (1^H) is true; otherwise, in uttering (3), the belief reporter communicates an instance of (3^H) in which Φ* is trivially satisfiable.

(P^H) is justified by the same appeals to the Maxim of Quantity that were used in defense of (P): if a speaker knows that *A* does not believe *p* under any mode of presentation, then, in uttering '*A* doesn't believe *p*' he will mislead if, for some contextually definite, nontrivial style Φ*, he asserts that there is *no m* meeting Φ* under which *A* believes *p*. The reason is obvious: if he does assert this, he must intend, or believe it highly likely, that his audience infer that, for some contextually definite, possibly trivial style Γ*, there is some *m'* meeting Γ* such that *A* does believe *p* under *m'*. Since he wants to avoid giving rise to false beliefs, he asserts that for some trivially constrained *m*, it is not the case that *A* believes *p* under *m*. This allows him to say something true (that *A* does not bear the *believes* relation to *p*) without endorsing a false belief (that *A* does bear the *believes* relation to *p*).

Although quite attractive, this proposal nonetheless raises worries of its own. Recall our central example, introduced in section VI, according to which Lois meets Kent before she learns of Superman's existence—at *t*. In uttering (3) at *t* there seems to be no particular reason to place any substantive constraints on *how* the referred-to proposition is presented to Lois, since, after all, it is mutually understood between speaker and hearer that there is *no guise* under which she assents to it. Such an utterance must either constrain the style vacuously (for example, imposing a condition of self-identity), or involve no style reference at all. If so, our speaker's utterance of (1) at *t'*, after Lois has learned of Superman's ability to fly, will contradict his previous utterance of (3). On the other hand, if we suppose that the speaker at *t* substantively constrains the style in some way or other—perhaps he asserts that Lois does not believe **s** under a mild-mannered-reporter-presenting guise—then he has violated the Maxim of Quantity. By being more informative than is strictly necessary, he would present himself as believing that, for all he knows, there *is* some mild-mannered-reporter-presenting guise under which Lois does believe **s**. But this is absurd.

Thus, if HIT is to save semantic appearances and avoid contradiction, it must have our well-informed and well-intentioned belief reporter generating a false implicature and thus intentionally misleading his audience. That it is hardly worth saving semantic appearances at such a cost is perhaps too obvious to mention.

CONCLUSION

The central example of this paper is a rather simple variation on a well-known scenario, but it exposes a deep problem for the two accounts of belief ascription currently dominating the market—a puzzle that remains even after Saul Kripke's Pierre case has been resolved.[43] It is beyond the scope of this paper to determine whether any other, less popular, theories of belief ascription can avoid the puzzle I have described. For that matter, I cannot here address the question whether any case can be made for suspending or restricting the Maxim of Quantity (or Levinson's Q-principle), so as to allow our hypothetical speaker to be more informative than the described transaction seems to require. My concern has been simply to expose a tension between each of the aforementioned proposals and a seemingly inescapable fact about the norms of conversation—though I submit that abandoning one or the other account of belief is preferable to challenging the maxim in question.

GARY OSTERTAG

Nassau Community College

---

[43] Kripke, "A Puzzle about Belief," in Avishai Margalit, ed., *Meaning and Use* (Boston: Reidel, 1979), pp. 239–83; reprinted in Salmon and Soames, eds., *Propositions and Attitudes* (New York: Oxford, 1988), pp. 102–48.

BOOK REVIEWS

The Discovery of Freedom in Ancient Greece. KURT RAAFLAUB. Translated by RENATE FRANCISCONO; revised by the author. Chicago: University of Chicago Press, 2004. xi + 420 p. Cloth $55.00.

The ancient Greeks discovered the concept of political freedom; and ever since, philosophers have wondered whether it might not have been better left in oblivion. Plato scathingly mocked the ideal of *eleutheria* or freedom for its disregard both for the good and for established custom and law.[1] Hobbes, while approving of the Greek notion of the unrestricted freedom of states, claimed that the West had mistakenly applied this notion to individuals as well. Hence, he claimed, men learned from the Greek and Latin authors "the habit of favoring tumults, and licentious controlling of their sovereigns...with the effusion of so much blood that I think I may truly say: there was never anything so dearly bought, as these western parts have bought the learning of Greek and Latin tongues."[2] Nor has opposition to Greek freedom been exclusive to monarchists. Benjamin Constant famously warned against favoring what he called "the liberty of the ancients"—or share in political power without protection from it—over the "liberty of the moderns," or freedom from collective interference in individual life.[3]

In this way, philosophers and historians of ideas have raised two sorts of questions, the first about what sort of freedom it was that the Greeks valued; and the second, whether or not that sort of freedom is a good thing. The pursuit of the first question, and through it the second, should be greatly aided by the new edition and translation from German of Kurt Raaflaub's 1985 monograph *The Discovery of Freedom in Ancient Greece. Discovery* is without question a magnificent book. It provides an exhaustive survey of the idea of freedom from Homer to the end of the fifth century B.C., as well as numerous plausible and interesting proposals for understanding its original development and its change over time. Raaflaub's careful attention to period and context gives the book a subtlety and richness of detail that match its broadness of scope. It is true that this same subtlety can

---

[1] *Republic* 557a–564a; *Laws* 701a–d.

[2] *Leviathan* 21.9.

[3] *The Liberty of the Ancients Compared with That of the Moderns.*

0022-362X/05/0211/594–601

make it difficult to find general patterns in what the Greeks thought about freedom; and in addition, those interested in philosophical issues may find themselves frustrated with Raaflaub's failure to signal emphatically key conceptual distinctions. The book's theses about the historical conditions surrounding the concept of freedom are much more clear than its theses about what the concept actually means and how its meaning changes over time. For this reason, reading the book can feel a bit like investigating the nature of an obscure projectile by tracing its path through the air. All the same, the book's lack of clarity on conceptual matters is minor when it is not superficial, and such general patterns as there are in Raaflaub's picture of Greek freedom are well worth the time and effort it takes to untangle them.

Raaflaub's general project is to explain the emergence of a specifically political concept of freedom and to trace developments in its use from the time of Homer up until the end of the fifth century and the beginnings of Greek political theory. Major conceptual changes are plausibly explained by major historical events, chiefly the Persian invasion of Greece in the early fifth century and the growth of the Athenian empire in the aftermath of the Persian wars. The ideal of freedom in these contexts is the freedom of whole city-states, not of individuals. The final stage in Raaflaub's account of the concept of freedom is the emergence of a specifically democratic conception of freedom, which he dates to the middle of the fifth century and the rivalry between Athens and Sparta. The democratic conception of freedom, Raaflaub argues, was at least in part an ideal for individuals as well. In what follows I will describe the key moments in Raaflaub's history of freedom that illustrate its collective character. Then I will discuss in detail Raaflaub's account of the democratic ideal of freedom and the freedom of individuals.

Raaflaub argues that the historical watershed that launches the concept of freedom into its central role in Greek political life is the invasion of mainland Greece by Persia early in the fifth century B.C. (chapter 3).[4] This invasion presented an unprecedented threat of foreign domination, sparking a newfound concern for the independence and self-determination of Greek city-states. At the same time, the prolonged encounter with a political culture the Greeks saw as tyrannical and slavish led them to value anew the constitutions under which they lived: structured, law-governed, and providing citizens some degree of participation in their governance. Thus the Persian wars inspired concern for both the independence of polises from external control (what Raaflaub calls "external freedom") as well as for internal political structures that discouraged or prevented tyranny (what Raaflaub calls "internal freedom"). The joint development of the ideal of self-determination with the ideal of freedom from tyranny was furthered by the Persian sponsorship of local tyrants and their use of them as proxies in those cities they did subjugate.

---

[4] In what follows I reference chapters by topic, omitting chapter 1 (on methodological issues), chapter 2 (on the concept of freedom before the Persian wars), and much of chapter 7 (summary and final considerations).

The second great watershed in the development of the political concept of freedom is the growth of the Athenian empire in the aftermath of the Persian wars (chapter 4). The Athenian empire influenced the development of the concept of freedom first of all by forcing more clear and specific reflection on the nature of self-determination. As a result of the Athenian allies' extended negotiations and failures to negotiate their independence, self-determination was no longer seen simply "from the outside," as freedom from external authority. Rather, a new term, *autonomia*, was coined to describe self-determination as a set of internal conditions such as sovereignty over taxation or judicial matters. Raaflaub elegantly argues that *eleutheria* and *autonomia* should be understood as different perspectives on similar conditions, the first a negative term concerning the absence of outside domination, the second a positive term indicating conditions of self-determination. In this way one can have *autonomia* while under outside rule (that is, without *eleutheria*), but the two words can also describe identical objective conditions.

The Athenian empire also changed the concept of freedom by inspiring or pressuring the Athenians to give some account of the ideal of freedom that was compatible with its subjugation of other Greek states (chapter 5). The result is what Raaflaub calls the Athenian doctrine of absolute freedom, the identification of freedom with unrestricted power. Power over others is not only a condition for freedom, but constitutive of it, an idea neatly summarized in the words of Aeschylus: "No one is free but Zeus."[5] For Raaflaub, the doctrine of absolute freedom does not only concern interstate relations, but is rather a comprehensive political concept referring to internal political conditions as well. Imperialism, for the Athenians, was thus conceptually closely tied with democracy: the Athenian *dēmos* or citizenry did not only rule itself, but also ruled over others.[6] In other words, unlimited power begins at home.

---

[5] *Prometheus Bound* 50, cited by Raaflaub on p. 187.

[6] The Greek word *dēmos* can refer both to the whole citizenry and to a part of that citizenry, the poor or "the many." I discuss the second, pejorative use below.

Raaflaub's last major topic is the relation between freedom and democracy (chapter 6). By the fourth century, in Plato and Aristotle, we find freedom closely associated with democracy as opposed to other types of constitution. Raaflaub argues that the connection of freedom with democracy appears late in democracy's development. (By contrast, the concept of equality was closely tied to democracy from the beginning.) The idea that democracies are freer than other nontyrannical types of government appears first in the 430s in the writings of Herodotus and the Old Oligarch (the author of the antidemocratic tract *Constitution of the Athenians*). The association of democracy with freedom thus seems to coincide with discussions of the contrast between democracy and oligarchy, most particularly contrasts between Athens and Sparta. The specifically democratic ideal of freedom that emerged was not just freedom for the demos or citizenry as a whole (as entailed by the Athenian doctrine of absolute freedom), but it involved liberty for individuals as well.

Before I turn to the Athenian ideal of individual freedom in detail, it is worth emphasizing how surprising an ideal of individual freedom appears in this context. The most striking general pattern in the Greek concept of freedom as Raaflaub describes it is its overwhelmingly collective character. This is most obvious when the freedom in question is the freedom of a polis from Persian domination or Athenian hegemony. More surprisingly, a city's freedom from tyranny, a matter of its internal political arrangements, is also a collective form of freedom: the freedom of a whole community from domination by a single man. Likewise, even the democratic concept of *isonomia*, equal distribution of political power combined with equality under the law, is not primarily a matter of the protection of individual liberties. Rather, it is a way of securing the liberty of the whole community; as Raaflaub puts it elsewhere, "individual political rights, important though they are, count mainly as parts of a system that can represent and guarantee freedom only as a whole."[7] In other words, all are counted equally not so much to give power to each, as to give power to *all.*

On the other hand, Raaflaub argues, once democracy and oligarchy are seen as ideological rivals, the notion of democratic freedom develops and expands to include a more robust kind of individual freedom. While it is undeniable that democratic freedom for the Athenians did include freedom for individuals, I fear that Raaflaub exaggerates its importance and blurs some of its key details.

---

[7] "Democracy, Oligarchy, and the Concept of the 'Free Citizen' in Late Fifth-Century Athens," *Political Theory*, XI (1983): 517–44, here p. 521.

A constitution, Raaflaub reminds us, was more than just a political system for the Greeks; it entailed a comprehensive way of life, political, social, and individual. Given this, it is not surprising that we find, in the period of heightened awareness of the differences between Athens and Sparta, passages that indicate differences in lifestyle between the two cities. Most notable here is Pericles's "Funeral Oration" in Thucydides, where Pericles praises not only the freedom and power of the Athenian demos as a whole, but the freedom of their way of life as well. Pericles hence extols their tolerance for their neighbor "if he enjoys himself in his own way"[8] and boasts of their individual virtue and self-sufficiency. These passages are echoed later in the book when the general Nicias, in the face of battle, reminds his men that all in Athens "had the liberty to live their lives in their own way."[9] Thus Athenian freedom for individuals meant a free manner of life as well as a share in political power.[10]

It is an interesting question, and one that Raaflaub does not raise directly, just how wide the scope is of the free manner of life that Thucydides and others describe, or how much freedom is actually entailed in such expressions as living "in one's own way" or "as one likes." It is worth stressing that the key historical contrast here is with Sparta and its highly restrictive and austere lifestyle. So whatever the actual scope of individual freedom in Athens, the bar in these contexts has been set rather low.

Raaflaub does point out a significant and nonobvious aspect of Athenian freedom: that it provides key opportunities for social advancement and social status.[11] In this way there is at least one sense that Athenian democracy provided more options for individuals: the lower social strata faced fewer social-class-based restrictions on what they could say or do. As the Old Oligarch notes, the noncitizen slaves live more freely in Athens;[12] and Plato complains in *Republic* VIII that even the animals in a democracy do not seem to know their proper station.[13]

---

[8] Thucydides II.37.

[9] Thucydides VII.69.

[10] However, Pericles also means to emphasize the contribution the excellence of individuals makes to the whole; and it seems to me plausible that the whole here is of primary importance.

[11] See also Raaflaub, "Democracy, Power, and Imperialism in Fifth-Century Athens," in J. Peter Euben, John R. Wallach, and Josiah Ober, eds., *Athenian Political Thought and the Reconstruction of American Democracy* (Ithaca: Cornell, 1994), pp. 103–46, especially pp. 138–46.

[12] Pseudo-Xenophon, *Constitution of the Athenians* 10–12; discussed by Raaflaub on *Discovery*, p. 222.

[13] *Republic* 562b.

All the same, it still seems reasonable to ask, given the overwhelming emphasis on the collective and the community in Athenian life, how significant the ideal of individual freedom was. Was it a mere social side effect of the collective freedom that receives so much more emphasis? Or did it play any role in Athenian justification of democracy, as Raaflaub argues it did?

For Raaflaub, the coinage of the word *parrhēsia* (freedom of speech) in the last third of the fifth century is key to understanding the development of the democratic notion of freedom. *Parrhēsia* constitutes "the free person's unrestricted right to express his opinion" and so is required for the freeman to "achieve full self-realization" (223). As such, *parrhēsia,* "more than anything else, constituted democratic freedom" (225). Furthermore, Raaflaub suggests, it has a key individualistic component.[14]

I find Raaflaub's discussion here puzzling and problematic. It is clear that *parrhēsia* often means to "speak freely," that is, to speak frankly and without fear of reprisal. It is also true, as Raaflaub perceptively shows, that the privilege of speaking freely was identified with the status of citizenship; to be without it, as several passages in Euripides show, is to be like a slave (223). But I am not sure that it follows from these two features of the word that it constitutes an unrestricted right to express one's opinion. The freedom connected with the word seems to have quite a narrow meaning: a freedom to speak the truth without reprisal; or somewhat more broadly, the privilege of giving "good advice" to one's city.[15] This is surely narrower than the freedom to say whatever one wants.[16] Likewise, the identification of the right of *parrhēsia* with citizenship and its denial with slavery does not seem enough to draw Raaflaub's conclusions. A slave has a restricted sphere of action, but more than that, he is a foreigner, an alien. The notion of freedom of speech thus uncomfortably straddles the idea of a narrow option (speaking without reprisal) and what Raaflaub calls self-fulfillment (participation in one's community through citizenship). I do not know that separately or together they can amount to an unrestricted right to express one's opinion.

---

[14] As on, for example, p. 276.

[15] As in Euripides, *Suppliants* 438–41, discussed on *Discovery,* p. 228.

[16] There are instances of *parrhēsia* in Isocrates (for example, *Areopagiticus* 20.7, *Busiris* 40.3) where it is used to say something shameful or against *nomos.* But this will not be enough to widen the scope of the relevant freedom: after all, it could be that Isocrates is describing something considered "truth" or "good advice" from another perspective. For similar reasons it seems unwise to look to democracy's critics for the scope of *eleutheria.*

Perhaps the deeper source of my discomfort here is that Raaflaub comes close to calling *parrhēsia* a right of self-expression. But this cannot mean self-expression in our sense, as the expression of a unique and individual point of view or set of values. Athenian democracy does not seem to be based on pluralistic ideals; it does not seem to value the inclusion of a plurality of irreducibly individual perspectives or Rawlsian conceptions of the good.[17] Perhaps Raaflaub means rather that *parrhēsia,* because it is equivalent to citizenship, is means of self-fulfillment and self-development. If so, this notion of individual democratic freedom may still be what Constant thought it was: self-fulfillment by participating in a whole larger than oneself.[18] Even so, I do not see how the use of *parrhēsia* illuminates this notion of self-fulfillment, nor does Raaflaub make it clear that this is the sense of individual freedom that he means.

In summary, I am not convinced that individual freedom was a significant political ideal for the Athenians. It seems to me questionable to put much weight on the individual Athenian freedom described by Thucydides, as these passages can be more modestly explained by Athenian pride in the cultural differences between themselves and their enemy in wartime. Nor do I see a convincing case that the concept of *parrhēsia* shows the importance of individual freedom in justifying democracy. Lastly, although overly "free" social and individual behavior is prominent in anti-democratic writing from the Old Oligarch through Plato and Aristotle, this alone would not tell us how Athenian democrats conceived of such behavior. Nor is it even clear how important freedom in individual lifestyle is in the ancient critiques of democracy. It is certainly more common for the critics to challenge the idea that the ruling *dēmos* constitutes the whole city: for them, democracy is the rule of a single faction, the poor. (The word *dēmos* was used in Greek both to refer to the citizenry as a whole and to refer to a particular faction, the poor or "the rabble.")

The emphasis Plato seems to put on individual liberty in his critique of democracy in *Republic* VIII may be misleading on this point. After all, the central question of the *Republic* is the choice of lives, and individual character-types are a key part of the way that question is framed. It is true that Plato is also interested in ways in which the concept of liberty may inadvertently justify tyranny, and one key step in that justification may be the application of a collective value to an individual. The simplest argument of that form may be that the Athenian doctrine of unrestricted freedom for the polis applied to individuals justifies the unconstrained rule of the strongest individual.[19] But Plato here is intentionally turning Athenian ideals inside out and trying to show that they mean something they were never intended to mean.

---

[17] If this is right, it may also partly explain the tension Josiah Ober describes between the democratic ideal of freedom and the ideal of consensus or *homonoia* (*Mass and Elite in Democratic Athens* (Princeton: University Press, 1989), pp. 295–99).

[18] C.f. also Isaiah Berlin's notion of positive freedom as described in *Two Concepts of Liberty* (New York: Oxford, 1958).

In Raaflaub's conclusion, he discusses the Greek idea that the relations within a polis were a microcosm of the relations between polises, as an explanation for the pervasiveness of freedom within the democratic polis as well as without. But the application of collective ideals to individuals and vice versa did not happen in an unrestricted fashion; otherwise, the "freedom as power" doctrine could have justified tyranny. Rather, for the Athenians, the collective group of citizens is the primary bearer of freedom; individuals share in this primarily by directly sharing in its power. It may well be of secondary importance to them that individuals are also free by having a less restrictive lifestyle than they might elsewhere.

I should note that these tentative critical remarks are partly influenced by the strength of Raaflaub's own account of the Greek ideal of collective freedom. By that it should be all the more clear that my criticisms are not meant diminish the accomplishment of this vast and thought-provoking book. It provides an unrivaled opportunity to better understand the most dramatic and controversial political legacy of the Greeks.

ZENA HITZ

Auburn University

---

[19] I am not sure Plato ever makes this argument explicitly, but something along these lines is clearly suggested in the *Gorgias*. Two excellent recent discussions of the connection between democratic and tyrannical ideals in the *Gorgias* are Rachel Barney, "Plato on the Two Faces of Rhetoric" (manuscript) and Rachana Kamtekar's "The Profession of Friendship: Callicles, Democratic Politics, and Rhetorical Education in Plato's *Gorgias*" (forthcoming in *Ancient Philosophy*).

NEW BOOKS

NUZZO, ANGELICA: *Kant and the Unity of Reason.* West Lafayette, IN: Purdue, 2005. xvi, 398 p. Cloth $62.95, paper $32.95.

ODDIE, GRAHAM: *Value, Reality, and Desire.* New York: Oxford, 2005. xvi, 252 p. Cloth $65.00.

OLIVER, KELLY: *The Colonization of Psychic Space: A Psychoanalytic Social Theory of Oppression.* Minneapolis: Minnesota UP, 2004. xxiv, 247 p. Cloth $59.95, paper $19.95.

OLSARETTI, SERENA: *Liberty, Desert, and the Market: A Philosophical Study.* New York: Cambridge, 2004. viii, 184 p. Cloth $70.00.

OMNÈS, ROLAND: *Converging Realities: Toward a Common Philosophy of Physics and Mathematics.* Princeton: University Press, 2005. xxvii, 264 p. Cloth $29.95.

OPPEL, FRANCES NESBITT: *Nietzsche on Gender.* Charlottesville: Virginia UP, 2005. xiv, 221 p. Cloth $49.50, paper $18.50.

OPPENHEIM, FRANK M., S.J.: *Reverence for the Relations of Life.* Notre Dame: University Press, 2005. xxi, 498 p. Cloth $68.00.

OSBORNE, THOMAS M.: *Love of Self and Love of God in Thirteenth-Century Ethics.* Notre Dame: University Press, 2005. 325 p. Cloth $60.00, paper $30.00.

PAPPAS, NICKOLAS: *The Nietzsche Disappointment.* Lanham, MD: Rowman and Littlefield, 2005. xvii, 266 p. Cloth $75.00, paper $27.95.

PENSKY, MAX: *Globalizing Critical Theory.* Lanham, MD: Rowman and Littlefield, 2005. vi, 254 p. Cloth $75.00, paper $27.95.

PEPERZAK, ADRIAAN THEODOOR: *Philosophy: Between Faith and Theology.* Notre Dame: University Press, 2005. xiv, 216 p. Cloth $46.00, paper $22.00.

PETERS, JOHN DURHAM: *Courting the Abyss: Free Speech and the Liberal Tradition.* Chicago: University Press, 2005. viii, 309 p. Cloth $29.00.

PIPPIN, ROBERT P.: *The Persistence of Subjectivity: On the Kantian Aftermath.* New York: Cambridge, 2005. viii, 369 p. Cloth $75.00, paper $28.99.

PRED, RALPH: *Onflow: Dynamics of Consciousness and Experience.* Cambridge: MIT, 2005. xii, 348 p. Cloth $38.00.

PRICE, TERRY L. *Understanding Ethical Failures in Leadership.* New York: Cambridge, 2006. xiii, 224 p. Cloth $65.00, paper $24.99.

PRITCHARD, DUNCAN: *Epistemic Luck.* New York: Oxford, 2005. 290 p. Cloth $55.00.

RASMUSSEN, DOUGLAS B., and DOUGLAS J. DEN UYL: *Norms of Liberty: A Perfectionist Basis for Non-Perfectionist Politics.* University Park: Penn State UP, 2005. xviii, 358 p. Cloth $80.00, paper $25.00.

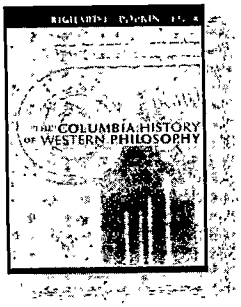
REIFF, MARK R.: *Punishment, Compensation, and Law: A Theory of Enforceability.* New York: Cambridge, 2005. 272 p. Cloth $70.00.

REISCH, GEORGE A.: *How the Cold War Transformed Philosophy of Science: To the Icy Slopes of Logic.* New York: Cambridge, 2005. xiv, 418 p. Cloth $70.00, paper $26.99.

REYDAMS-SCHILS, GRETCHEN: *The Roman Stoics: Self, Responsibility, and Affection.* Chicago: University Press, 2005. xi, 210 p. Cloth $35.00.

COLUMBIA, *Refer to us*

columbia.edu/cu/cup

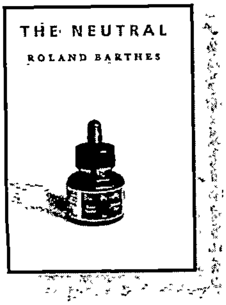


Columbia History of Western Philosophy

Edited by Richard H. Popkin

Lively, in-depth, and up-to-date historical analyses of all the key figures, schools, and movements of Western philosophy.

paper - $29.95



The Neutral

Roland Barthes

Translated by Rosalind Krauss & Denis Hollier

Roland Barthes's renowned and never-before-translated lectures pursue a central theme in his work: the quest for the neutral.

cloth - $29.50

Portrait of Jacques Derrida as a Young Jewish Saint

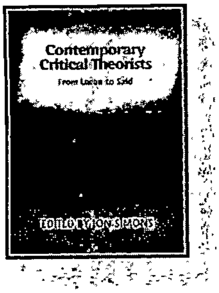
Hélène Cixous

Translated by Beverly Bie Brahic

"The book will have a special status within Derrida studies. . . . The most striking thing in Cixous's writing is the sense that something very confidential is being disclosed."
—*Jewish Quarterly*

paper - $19.95

AVAILABLE FROM EDINBURGH UNIVERSITY PRESS



Contemporary Critical Theorists

From Lacan to Said

Edited by Jon Simons

This authoritative guide introduces the key figures in contemporary critical theory.

paper - $29.00

You can now subscribe to the

THE JOURNAL OF
PHILOSOPHY

using

PayPal

New Rates January 2006

Institutional	$100
Individual	$45
Student/Retired	$20

To learn more about this easy new option, click on the PayPal icon at www.jphil.org

New From Routledge

Routledge Philosophers is a major new series of introductions to the great Western Philosophers. Each book places a major philosopher or thinker in historical context, explains and assesses their key arguments, and considers their legacy. Additional features include: a chronology of major dates and events, chapter summaries, annotated suggetions for further reading, and a glossary of technical terms. An ideal starting point for those new to philosophy, they are also essential reading for those interested in the subject at any level.

FREUD

Jonathan Lear, University of Chicago

"This book will be viewed by philosophers interested in psychoanalysis as a major contribution."

—Richard Rorty, Stanford University

"Jonathan Lear succeeds brilliantly in showing Freud to be philosophically significant–a philosophically ambitious, stimulating, enthusiastic, passionate and exciting book."

—Sebastian Gardner, University College London. UK

Pb • $22.95

HEGEL

Frederick Beiser, Syracuse University

"An impressive achievement, I have no doubt students will find it very useful, and that it will be widely accepted as a teaching text: it is written in a clear and accessible manner; it covers the right topics to the right level; it engages with a wide range of Hegel's works; it is critical, while also being sympathetic; and it deals authoritatively with various matters of scholarship."

—Robert Stern, Sheffield University

Pb • $22.95

HOBBES

A.P Martinich, University of Texas at Austin

"This is an excellent book, well suited to the aims of the Routledge Philosophically *series. It is clearly and accessibly written, comprehensive, up to date on current scholarship, well organized, and often humorous. I think undergraduates will find the book both readable and enjoyable. Teachers will find it very helpful."*

—S A Lloyd, University of Southern California

Pb • $22.95

LOCKE

E.J. Lowe, University of Durham, UK

"A wonderful contribution to the understanding of Locke."

—Dan Kaufman, University of Colorado, Boulder

"An admirable introduction and written by a real philosopher with a real philosophical interest in Locke."

—Andrew Pyle, Bristol University, UK

Pb • $22.95

ROUSSEAU

Nicholas Dent, Birmingham University, UK

"A very fine piece of work indeed. In fact, I think it is the best general introduction to Rousseau's life and thought in English and succeeds brilliantly in conveying Rousseau's ideas in a sympathetic yet critical manner."

—Christopher Bertram, University of Bristol

Pb • $22.95

SCHOPENHAUER

Julian Young, University of Auckland, New Zealand

"This is a comprehensive, sensible and lucidly presented account of Schopenhauer's philosophy. The author succeeds in showing why Schopenhauer still deserves to be studied, but at the same time he is appropriately critical when discussing Schopenhauer's doctrines and arguments."

—Severin Schroeder, University of Oxford

Pb • $22.95

THE SHORTER ROUTLEDGE ENCYCLOPEDIA OF PHILOSOPHY

Edited by Edward Craig, University of Cambridge

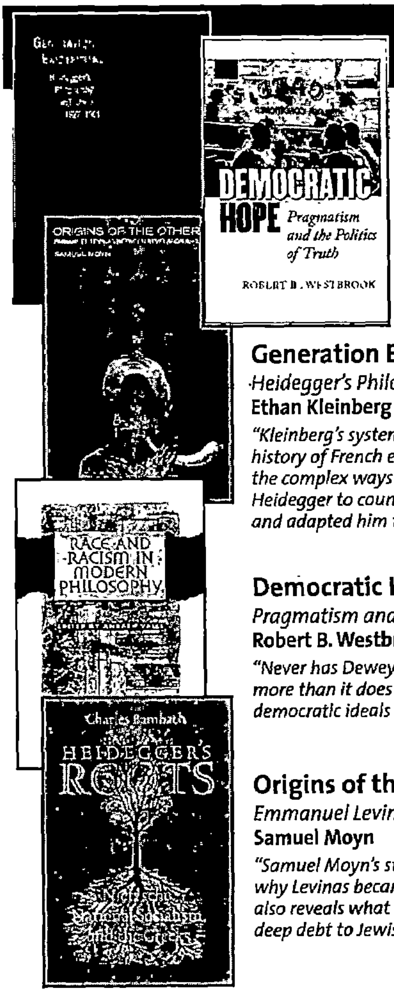
"Indispensable to all but the most solitary of philosophers; to historians of thought and culture; to students everywhere."

—George Steiner, *New York Times Book Review*

Hb • Only $45.00

Routledge
Taylor & Francis Group

1.800.634.7064 □ WWW.ROUTLEDGE-NY.COM



New Titles in Philosophy

CORNELL
UNIVERSITY
PRESS

Generation Existential
Heidegger's Philosophy in France, 1927–1961
Ethan Kleinberg

"Kleinberg's systematic overview of this crucial chapter in the history of French existentialism and its broader impact explores the complex ways in which French philosophers appropriated Heidegger to counter the French postwar intellectual establishment and adapted him to address their own specific priorities."

—Carolyn J. Dean, Brown University
$39.95 cloth

Democratic Hope
Pragmatism and the Politics of Truth
Robert B. Westbrook

"Never has Dewey's vision of the ethics of democracy mattered more than it does today, and no one illuminates American democratic ideals better than Westbrook does in this book."

—James T. Kloppenberg, Harvard University
$29.95 cloth

Origins of the Other
Emmanuel Levinas between Revelation and Ethics
Samuel Moyn

"Samuel Moyn's study of Levinas not only helps us understand why Levinas became such a revered figure in postwar France but also reveals what that reverence obscured—in particular, Levinas's deep debt to Jewish and, surprisingly, Christian theology."

—James T. Kloppenberg, Harvard University
$29.95 cloth

Race and Racism in Modern Philosophy
edited by Andrew Valls

"This collection raises the current debate over the extent to which racism has distorted our philosophical legacy to a new level of scholarship and sophistication."

—J. L. A. Garcia, Boston College, author of *The Heart of Racism*
$45.00 cloth, $21.95 paper

Heidegger's Roots
Nietzsche, National Socialism, and the Greeks
Charles Bambach

"Bambach's account. . . . places Heidegger's thought within a historical and sociopolitical perspective that will be essential to Heidegger scholars and those who see his work as relevant to contemporary hermeneutics."

—*Journal of the History of the Behavioral Sciences*
$24.95 paper

For a full list of all of our titles in Philosophy, including new books in our Ancient Commentators on Aristotle series, please visit

ww.cornellpress.cornell.edu

CAMBRIDGE

OUTSTANDING SCHOLARSHIP

Economic Compulsion and Christian Ethics

Albino Barrera

Albino Barrera argues that Christian thought on economic security offers an effective framework within which to address the consequences of economic compulsion. Such economic ripple effects of market operations have been virtually ignored in ethical discourse because they are generally accepted to be the very mechanisms that shape the market's much-touted allocative efficiency.

New Studies in Christian Ethics 24

$75.00. Hardback: 0-521-85341-9: 266pp

Leo Strauss and the Theological-Political Problem

Heinrich Meier

The theologico-political problem refers to the confrontation between the theological and the political alternative to philosophy as a way of life. In this study, Heinrich Meier clarifies the distinction between political theology and political philosophy and sheds new light on the unifying center of Strauss' philosophical work.

Modern European Philosophy

$60.00. Hardback: 0-521-85647-7: c.224pp

The New Hegelians

Politics and Philosophy in the Hegelian School

Edited by Douglas Moggach

The period leading up to the Revolutions of 1848 is a seminal moment in the history of political thought, demarcating the ideological currents and defining the problems of freedom and social cohesion, which are among the key issues of modern politics. This anthology offers new research on Hegel's followers in the 1830s and 1840s. Including essays by well-known philosophers, political scientists, and historians from Europe and North America, it pays special attention to questions of state power, the economy, poverty, and labor, as well as to a range of ideas about freedom.

$65.00*: Hardback: 0-521-85497-0: c.350pp

*Prices subject to change.

Eye for an Eye

William Ian Miller

Analyzing the law of the talion – an eye for an eye, tooth for a tooth – literally, William Ian Miller presents an original meditation on the concept of "pay back". Miller's unique theory of justice offers redemption via retaliation. It espouses the view that revenge is a highly structured phenomenon that requires a deep commitment to balance in order to get even in a strict but fair manner.

$28.00*: Hardback: 0-521-85680-9: c.304pp

Genetics and Christian Ethics

Celia Deane-Drummond

This volume applies the principles of Christian ethics in examining recent significant developments in the science of genetics. It considers ethical issues arising out of specific practices in human genetics, including genetic screening, gene patenting, gene therapy and genetic counseling as well as feminist concerns.

New Studies in Christian Ethics 25

$75.00: Hardback: 0-521-82943-7: 303pp

$29.99: Paperback: 0-521-53637-5

The Spiritual Dimension

Religion, Philosophy and Human Value

John Cottingham

Offering a new model for the philosophy of religion, John Cottingham combines emotional and intellectual aspects of our human experience, and embraces practical as well as theoretical concerns. Cottingham reveals how a religious worldview is best understood not as an isolated set of doctrines, but as intimately related to spiritual praxis and to the search for self-understanding and moral growth.

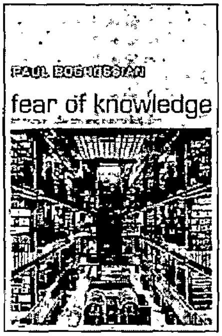
$70.00: Hardback: 0-521-84377-4: 198pp

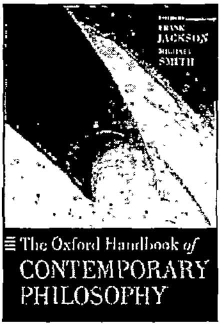
$24.99: Paperback: 0-521-60497-4

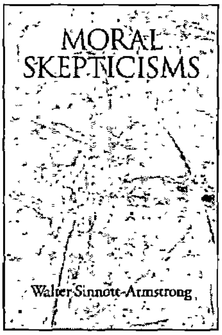
www.cambridge.org/us/philosophy/

CAMBRIDGE
UNIVERSITY PRESS

New From
OXFORD







Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

VISIBLE IDENTITIES

Race, Gender, and the Self

LINDA MARTÍN ALCOFF

In the heated debates over identity politics, few theorists have looked carefully at the conceptualizations of identity assumed by all sides. *Visible Identities* fills this gap. Drawing on both philosophical sources as well as theories and empirical studies in the social sciences, Martín Alcoff makes a strong case that identities are not like special interests, nor are they doomed to oppositional politics, nor do they inevitably lead to conformism, essentialism, or reductive approaches to judging others.

(Studies in Feminist Philosophy)
2005 paper $24.95 cloth $74.00

LOGICAL PLURALISM

J. C. BEALL and GREG RESTALL

In this book J.C. Beall and Greg Restall present and defend what they call *logical pluralism*, the view that there is more than one genuine deductive consequence relation, a position which has profound implications for many linguists as well as for philosophers. We should not search for one true logic, since there are many.

2005 paper $29.95 cloth $65.00

FEAR OF KNOWLEDGE

Against Relativism and Constructivism

PAUL BOGHOSSIAN

Paul Boghossian, in his long-awaited first book, sweeps away relativist claims that there is no such thing as objective truth or knowledge, but only truth or knowledge from a particular perspective. He demonstrates clearly that such claims don't even make sense. His short, lucid, witty book shows that philosophy provides rock-solid support for common sense against the relativists, and will prove provocative reading throughout the discipline and beyond.

2005 $29.95

THE OXFORD HANDBOOK OF CONTEMPORARY PHILOSOPHY

Edited by FRANK JACKSON and MICHAEL SMITH

Jackson and Smith, themselves two of the world's most eminent philosophers, have assembled more than thirty distinguished scholars to contribute incisive and up-to-date critical surveys of the principal areas of research. The coverage is broad, with sections devoted to moral philosophy, social and political philosophy, philosophy of mind and action, philosophy of language, metaphysics, epistemology, and philosophy of the sciences.

(Oxford Handbooks)
2005 $99.00

MORAL SKEPTICISMS

WALTER SINNOTT-ARMSTRONG

All contentious moral issues—from gay marriage to abortion and affirmative action—raise difficult questions about the justification of moral beliefs. How can we be justified in holding on to our own moral beliefs while recognizing that other intelligent people feel quite differently and that many moral beliefs are distorted by self-interest and by corrupt cultures? Sinnott-Armstrong provides an extensive overview of these difficult subjects, looking at a wide variety of questions, including: Are any moral beliefs true? Are any justified? What is justified belief?

2005 $49.95

Batla
13·6·06

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NUMBER 12
DECEMBER 2005

page
603 *Logical Relations between Pictures* Jan Westerhoff
624 *Against Fictional Realism* Anthony Everett
650 INDEX TO VOLUME CII

***Published by* The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Christopher Peacocke, Philip Pettit, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.
You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2006

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U.S. copyright law. All rights reserved. For more information about our copyright policies, please consult our January issue or contact us.

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc., MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027. Periodicals postage paid at New York, NY, and other mailing offices.
POSTMASTER: Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2005 by the Journal of Philosophy, Inc. ISSN 0022–362X

THE JOURNAL OF PHILOSOPHY
VOLUME CII, NO. 12, DECEMBER 2005

LOGICAL RELATIONS BETWEEN PICTURES

It is generally assumed that logical relations are relations between sentences, logical formulae, or propositions. All of these are linguistic objects; obviously in the case of sentences or formulae, plausibly so in the case of propositions if these are conceived of as collections of sentences "saying the same thing." The linguistic or descriptive mode is one major way of representing the world, but not the only (and perhaps not even the most important) one. Another major mode of representation is depiction. It represents not by sentences or formulae but by paintings, drawings, sketches, engravings, maps, diagrams, or photographs.

To the best of my knowledge, no one has ever inquired into the logical relations between depictions. This is peculiar, for assuming that logic is a general theory of certain fundamental relations between our representations of the world it is hard to see why it should only apply to one class of representations and not to another one. I think there are two main reasons for this peculiarity.

First of all it is often assumed that logical relations could only hold between objects with a transparent syntactic structure. If implication is seen to be a relation between two formulae, holding if it is possible to transform the first into the second using a fixed set of rules (as in a natural deduction system), then the restriction to descriptive forms of representation is obviously implied. Paintings, drawings, and so on do not have fixed syntactic structures; they are not composed from a fixed set of primitive symbols by a collection of formation rules.[1]

---

[1] Indeed Nelson Goodman takes this to be the fundamental distinction between linguistic and nonlinguistic systems of representation. See *Languages of Art* (Indianapolis: Hackett, 1976), pp. 225–32.

0022-362X/05/0212/603–23

Second, common opinion holds that logical relations are only applicable to truth-bearing items. If implication is semantically understood to be the relation holding between two items such that whenever the first is true, the second must be true as well, we can only consider extending this to depictions if we assume that these, like descriptions, can be true or false. There are a number of problems associated with this idea, as noted by Jerry Fodor.[2] When is a picture of a cat on a mat true? If there is some cat on some mat? Or does it have to be a fat cat, as in the picture? Or the very cat pictured? Does it have to be the very same mat? The difficulty of answering these questions does not let pictures appear as the most natural truth-bearers. But in this case there cannot be an implication relation between them, semantically understood.

Both of these assumptions—that implication relations only hold between syntactic items, and only between truth-bearers—are false. A perspicuous demonstration of this has been given in Arnold Koslow's development of a structuralist theory of logic.[3] He defines an implication structure as *any* collection of objects (which have to be neither syntactic nor truth-bearing) on which an implication relation obeying a set of Gentzen-style conditions can be established. Logically complex objects are then defined in terms of this relation. The conjunction of two objects *A* and *B*, for example, is taken to be the weakest object *C* such that it implies *A* and it implies *B*. (By 'weakest', we mean that if any other object also implies *A* and *B*, it will also imply *C*.)[4] It then remains to be demonstrated that the objects so defined really have the properties ascribed to logically complex objects such as conjunctions, disjunctions, negations, and so on.

One peculiarity of Koslow's system is that collections of objects are not usually logically closed. Frequently logical complexes of objects from the collection will fail to be included in it. For example a collection with only *A* and *B* in it will fail to contain a conjunction of the two if they are not mutually entailing. This is due to the fact that the structuralist theory defines the implication structure on a fixed set of objects; logical complexes are then identified with particular objects from this collection. This is no fundamental limitation, however.[5] The collections can always be expanded so as to include the required logical complexes.

---

[2] See Fodor, *The Language of Thought* (Cambridge: Harvard, 1975), pp. 174–94.

[3] See Koslow, *A Structuralist Theory of Logic* (New York: Cambridge, 1992), and "The Implicational Nature of Logic: A Structuralist Account," in Achille Varzi, ed., *The Nature of Logic* (Stanford: CSLI, 1999), pp. 111–55.

[4] This account cuts a number of corners. For the full details, the reader is referred to Koslow's *A Structuralist Theory of Logic.*

[5] In fact, it might be important to allow for the possibility that logical operations are not always defined for all items under consideration. See Dana Scott, "Background to

In the following, I want to use the structuralist theory of logic to develop an account of logical relations between pictures. I will describe an implication relation between pictures and argue that it obeys the conditions mentioned by Koslow. It is then possible to give precise definitions of conjunctions, disjunctions, negations, and so on, of pictures. It will also turn out that these logical operations are closely related to or even identical with basic cognitive operations we naturally employ when thinking about pictures. Before this can be done, however, it is necessary to discuss two preliminary matters: the nature of the type-token distinction in the case of pictures and the relationship between a picture and its parts.

I. PRELIMINARIES

I.1. Pictures and Paintings. Pictures differ from paintings as propositions differ from sentences. Paintings (as well as drawings, sketches, engravings, and the like) and sentences are tokens: spatiotemporally located physical objects. Different paintings can show the same picture, and different sentences can express the same proposition. Pictures and propositions are what particular sets of paintings or sentences have in common, they are therefore not tokens but types.[6] Relative to some set of conventionally defined criteria we regard certain paintings as showing the same picture, and certain sentences as expressing the same proposition. What these criteria are in each particular case is notoriously hard to define, and it is very likely that no analysis of the type-token relation in terms of necessary and sufficient conditions we give will ever be satisfactory. This is, however, not a problem the present discussion will address; we will simply take the type-token distinction for visual representations as primitive.

In discussing the logical relations between visual representations, it seems evident that we should concentrate on types (on pictures) rather than on tokens (on paintings). Although it would be foolish to expect an implication relation between visual representations to be in every respect similar to the implication relation between propositions familiar to us from logic, this is certainly a plausible point of depar-

---

Formalization," in Hughes Leblanc, ed., *Truth, Syntax, and Modality* (Amsterdam: North-Holland, 1973), pp. 244–73, and "Completeness and Axiomatizability in Many-valued Logic," in Leon Henkin, ed., *Proceedings of the Tarski Symposium* (Providence, RI: American Mathematical Society, 1974), pp. 411–35.

[6] Whether to conceive of pictures and propositions as abstract objects or in terms of naturalistically more acceptable constructions from paintings and sentences is of little consequence for the present discussion.

ture for investigations into implication relations with other kinds of relata. As such we want to conceive of an implication relation between visual representations as a relation between what various collections of such representations "showing the same thing" have in common, rather than as a relation between the physical objects which are the basis of the showing or representing. We will therefore look at logical relations between pictures, not between paintings.

Note that due to the greater abstractness of pictures, there are some questions concerning them which do not have definite answers, even though the same questions asked about a painting instantiating them do have such answers. If we consider Botticelli's 1489 painting of the *Annunciation* (the tempera-on-wood object in the Uffizi) we can ask, for example, what kind of red pigment was used for Mary's cloak, or what the diameter of her halo is. The same cannot be asked about the *picture* of Botticelli's 1489 *Annunciation*, as the examination of various instances of the picture (a copy painted in oil on canvas, a postcard, an image on a computer screen) delivers radically different answers. This phenomenon can also be observed if we speak just of the picture of the Annunciation, rather than of Botticelli's. Whereas it makes perfectly good sense to ask about either the painting or the picture of Botticelli's *Annunciation* whether the angel approaches Mary from the left or the right, the same question is meaningless when asked about the picture of the Annunciation as such. (In the majority of cases the angel comes from the right, but not always.) It can therefore happen that we have a precise idea of what a picture depicts (Mary being visited by an angel) without having a fixed opinion on the spatial arrangement of some of its constituents.[7]

It is interesting to note in this context that Daniel Dennett assumes that "the rules of images in general" (where photographs and paintings are subsumed under images) exclude the lack of specificity just indicated.[8] It may be the case that the rules of *paintings* in general forbid such underdetermination regarding color, size, composition of parts, and so on, but it appears to be inadvisable to extend this to images or pictures.[9] Of course one would want to assume this if, as in the case of Dennett, the plan is to use this "rule" to argue that mental representations are description-like, rather than image-like. Dennett argues that since an imagined tiger has an undetermined number of stripes, the representation of the tiger must be like a description (which is also not forced to specify the number of stripes) and not like an image (which has to depict the tiger with some definite number) (*op. cit.*, p. 55). I think, however, that if we consider pictures instead of paintings this allegedly crucial difference in determinacy between description- and image-like representation disappears. If there is *any* difference between descriptions and pictures, Botticelli's *Annunciation* belongs to the latter. Yet as we have just seen there are various questions about this picture which do not have definite answers. That Botticelli had to give the Virgin's halo some size in the painting he painted does not imply that in the picture he thereby produced, her halo also has a definite size. Descriptions and depictions, it turns out, are both forms of representation which admit of underdetermination.

---

[7] The same considerations apply when comparing a particular inscription of the fifth commandment, the proposition expressed by it, and an arbitrary proposition forbidding murder.

[8] See Dennett, "The Nature of Images and the Introspective Trap," in Ned Block, ed., *Imagery* (Cambridge: MIT, 1981), pp. 51–61, on p. 55.

[9] For an argument that in fact such underdetermination can be present in pictorial *tokens*, see Michael Tye, "Image Indeterminacy," in Naomi Eilan, ed., *Spatial Representation* (New York: Oxford, 1993), pp. 357–60.

I.2. Parts of Pictures: Subpictures, Constituents, and Parts. What are the parts of a picture? The question is not as innocuous as it sounds. The mereology of ordinary objects is well developed, but pictures are no ordinary objects. First of all, they are not spatiotemporal, as we just saw. Second, they are structured: they are not like a heap of grain or a puddle of water the identity of which is preserved under various rearrangements of their parts. Pictures have parts which are put together in a certain way: if we destroy the order of the parts the picture is gone. Third, pictures are representational. A lego toy, a molecule, a pattern of tiles all have structure, but they do not usually represent.

While mereology deals with objects in time and space, there is no *prima facie* impossibility in extending the 'part of' relation to other kinds of things.[10] Recent mereological investigations have also started to tackle the second peculiarity,[11] while the third still remains to be addressed.

Some pictures have parts which are pictures themselves. A picture of Napoleon with a hat has a part which is the picture of a hat. So much is uncontroversial. But what about parts of pictures which are not themselves pictures? Here we are faced with two alternatives. First, we can conceive of pictures as arrangements of atomic colored pixels of arbitrarily small size on a plane.[12] Then any pixel and any fusion of an arbitrary collection of pixels from the picture of Napoleon's hat will be a part of the picture, in the same way as each hydrogen and each oxygen atom contained in it, as well as any arbitrary fusion of such atoms will be a part of a given puddle of water. Second, we may think of the hue, saturation, and brightness of the color of Napoleon's hat, its outline, the shading, and so on, as parts of the picture of the hat. I think the second conception is to be preferred to the first. Here is why.

---

[10] For example, one might regard the prime factors of a number as its parts. An application of mereology to such categories as properties or states of affairs can be found in Uwe Meixner, *Axiomatic Formal Ontology* (Boston: Kluwer, 1997).

[11] See Peter Simons, *Parts: A Study in Ontology* (New York: Oxford, 1987), p. 324.

[12] Note that this conception does not just confuse pictures with paintings. The picture would be taken to consist of pixels in the same way in which a water molecule

Consider the sense in which states of affairs can be taken to have parts. It is straightforward to argue that the state of affairs that John loves Becca has John as a part. But it is equally straightforward to argue (by the transitivity of parthood) that John's brain is part of the state of affairs that John loves Becca. Metaphysicians generally agree that there is an important difference between these two kinds of parts, a difference which is sometimes expressed by calling the first but not the second kind "constituents" of states of affairs. It is important to see what the difference between these two kinds is. A state of a affairs is a part of the world we pick out via a certain linguistic description ('John loves Becca', 'John liebt Becca', *Ljb*, and so on) and to which we ascribe a certain structure mirroring the form of the part of language we use to pick it out (the standard analysis is that the state of affairs that John loves Becca consists of two individuals and a dyadic relation). The constituents of the state of affairs are now precisely those parts which we use in conceptualizing this bit of the world; the elements which we take to correspond to the bit of language we use to identify it. But the mere parts (John's brain as opposed to John) are just *any* parts of that particular bit of the world we happen to be talking about, whether they take part in our conceptualization or not. In analyzing a state of affairs we therefore have to concentrate on its constituents, on the parts we actually use in thinking about this subsection of the world, rather than on any arbitrary collection of parts it happens to have. States of affairs are parts of the world which the human mind has sliced, and in investigating them we have to concentrate on the cuts the mind has made, rather than on any other fissures we may otherwise observe in the material.

A very similar situation confronts us in the case of pictures. These too are human creations. As such the outline, saturation, brightness,

---

consists of hydrogen and oxygen atoms. Different water molecules will have quantitatively distinct atoms as parts, but they will all belong to the same type of atom. In the same way the pixels, the parts of a picture can be different physical objects in the different physical realizations of a picture (they can be fragments of paint, drops of ink, or pieces of colored glass) but still all belong to the same type of pixel. A pixel, like a particular kind of molecule, is no token but a type.

hue, and so on, of a picture are more important in investigating them than arbitrary collections of pixels which happen to be parts of the picture. A picture is a representation which has been constructed by combining a certain outline with certain colors, certain shadings, and so on, rather than a collection of pixels which has been put together bit by bit. A collection of pixels from a picture of Napoleon's hat tells us no more about its role as a representation than the part of the state of affairs that is John's brain tells us about its role as a part of the world picked out by a particular linguistic description.

In the following investigation, we will therefore concentrate on constituents of pictures, rather than on their parts. Neither constituents nor parts are pictures themselves; if we want to refer to pictures which are contained in pictures we use the term *subpicture.* Note that constituents of pictures—unlike parts or subpictures—are not objects which can exist on their own: we cannot present an outline without a color, or a color without a shape. The constituents of a picture are ontologically dependent on one another.[13] Furthermore, as the reader is invited to check, all three kinds of parts of pictures are subject to the standard mereological axioms; the subpicture-, the constituent-, and the part-relation are reflexive, transitive, and antisymmetric.

One complication we should note, however, is that collections of pixels from a picture can fail to be either a part or a subpicture. They fail to be a part because they are a picture themselves, and they fail to be a subpicture because they are not contained in the picture.

To see how this can be the case consider the closely related case of sentences. Not every part of a sentence (string of symbols contained within it) which is a meaningful expression is also a constituent of that sentence. It must also be a meaningful expression *which is constitutive of the meaning of the entire sentence.* For example, in the sentence 'to you fall the rewards' the part 'you fall' is meaningful, but its meaning is not part of the meaning of the whole sentence and thus not a constituent.[14] Similarly, to use an example of Stephen Pinker's, the expression 'sex with Dick Cavett' is a constituent of the sentence 'the TV show discussed sex with Dick Cavett' only on one of its two possible readings.[15]

---

[13] The same is true of constituents of states of affairs: individuals depend on properties they instantiate, properties depend on individuals as instantiators. They are abstractions from states of affairs, but not things which could exist independently, outside the context of a state of affairs.

[14] This example is from Elliott Sober, "Mental Representation," *Synthese,* XXXIII (1976): 101–48, on p. 122.

[15] See Pinker, *The Language Instinct: The New Science of Language and Mind* (New York: Penguin, 1994), pp. 102–03.

A similar situation can arise with pictures by using clever cropping. In this case, a part of the picture is cut out which is indeed a picture itself, but none which could be discerned in the original picture with its original surroundings (and is therefore not a constituent). It is no more constitutive of the first picture than 'sex with Dick Cavett' is constitutive of the above sentence where 'sex' is the entire noun phrase and 'with Dick Cavett' a prepositional phrase (as opposed to the typographically identical one where both 'sex' and 'with Dick Cavett' are constituents of the noun phrase).

II. IMPLICATION RELATIONS BETWEEN PICTURES

There is a natural way of conceiving of an implication relation between pictures in terms of the relation of pictures and their subpictures. We might argue that in the case of two pictures, one of which shows a still life with fruits, and the other only a peach from this still life, the former implies the latter. More generally we could say that a picture implies all its subpictures.[16] Elliott Sober in his treatment of logical operations on picture agrees and claims that "the pictorial analogue of implications is containment" (*op. cit.*, p. 122). He also argues that containment is the *only* pictorial counterpart of implication. If this was indeed the case, there would not be much of interest in the study of implication relations between pictures. There would only ever appear one name of a picture on the left of any pictorial implication sign as only single pictures could imply anything individually, but never a group of them collectively. Moreover, implication would typically be unidirectional, as the item on the left of the implication-sign would contain more information than the one on the right.[17]

Neither of these two characteristics (which are independent of one another) is true of the familiar implication relation between propositions. Virtually all interesting implications have more than one premise, and while many are of the "information-losing" variety, in many other cases *implicans* and *implicatum* are equivalent.

---

[16] Note that this is an important respect in which pictures differ from sentences: 'The book is red or the book is heavy' implies neither of its two proper subsentences.

[17] Apart from the case of a picture being its own subpicture, which is always guaranteed by reflexivity. Note that there is also the case of a picture containing a near identical subpicture, like the label of a cheese showing a monk eating cheese on the label of which there is a monk eating cheese, and so on. Assuming finite complexity of pictures this regress has to stop somewhere (the label on the last cheese is just an array of pixels) and so this is *not* a case of a picture containing itself as a subpicture. If there are, say, five iterations then the largest picture will contain a part with only four iterations, which is obviously not identical with the largest picture.

Examples of each type can be conveniently summarized in the following table:

	ONE PREMISE		MULTIPLE PREMISES	
UNIDIRECTIONAL	$A \vdash A \vee B$	[1]	$A \rightarrow B, A \vdash B$	[3]
BIDIRECTIONAL	$A \rightarrow B \vdash \neg B \rightarrow \neg A$	[4]	$A, B \vdash A \wedge B$	[2]

If the pictorial analogue of implication was indeed containment, all implications between pictures would be of the one-premise unidirectional variety (type [1]). This, however, is not the case. All four possibilities given in the above table have pictorial analogues.

Let us first of all look at pictorial implication with multiple premises. For an example of the bidirectional case (type [2]) consider the following. Assume there is a fresco at the end of a room which we are only allowed to view from a certain distance. Unfortunately from our point of view there are always two columns blocking our view of parts of the fresco. We can move around the room to see any part of the fresco, but we can never observe the whole fresco without the columns. Now it is clear that we can piece together the appearance of the complete fresco in our mind from the parts we were able to observe. We could do a similar thing with photos we have taken from different points of view: cut out the columns and then arrange the fragments in such a way that we get a continuous picture.

I want to argue that the parts of the fresco thus observed collectively imply it. Not only will a picture imply any of its subpictures, a collection of pictures will also imply one which is the result of putting all of them together in a certain way. Drawing implications is not just reducing the information present in the premises ('The book is red and square', therefore 'The book is red'), but equally combining it into a compound ('The book is red', 'The book is square', therefore 'The book is red and square').

Now take a case with multiple premises which is unidirectional (type [3]). In this case two pictures imply a third one which contains less information than the two of them together. Assume we have two pictures of the same statue taken from two different positions, *A* and *B*. In some cases it is possible to work out what a view of the statue from some position *C* between *A* and *B* would look like, solely on the basis of the two pictures from positions *A* and *B*. Then we would also want to say that the picture from *C* is implied by the pictures from *A* and *B*. It would then contain less informational content than both the pictures which imply it put together.

Considering this example shows that it is unhelpful to conceive of pictorial implication in terms of "cut and paste" operations. These can account for unidirectional implications with one premise (where a subpicture is "cut out") and bidirectional implications with multiple premises (where the conclusion is "pasted together" from the premises). But "triangulating" a picture from two other ones in the way just sketched is not covered by such operations, nor are cases of bidirectional pictorial implications with one premise (type [4]). The most straightforward (if slightly boring) example is just the relation every picture bears to itself (the pictorial equivalent of $A \vdash A$). The fact that there are no more interesting examples of bidirectional pictorial implications with one premise (something like pictorial equivalents of $A \rightarrow B \vdash \neg B \rightarrow \neg A$) is due to the analogy between pictures and propositions described above. Pictures are collections of visual tokens which show the same, propositions are collections of linguistic tokens which say the same. Logical formulae "say the same" if they are logically equivalent, so $A \rightarrow B$ and $\neg B \rightarrow \neg A$ express the same proposition. Had we constructed the above table of different kinds of implication with propositions instead of logical formulae it would have been immediately obvious that the only example of a bidirectional implication relations with one premise would indeed have been of the form 'A implies A'.

I have argued that each of the four kinds of implication given in the above table have a pictorial analogue. I also want to claim that four examples I have described (subpicture implication, the fresco case, the statue case, and the case of pictures implying themselves) all involve *a single pictorial implication relation*. What is my argument for this? The most elegant argument would be to reduce the four examples to one, by defining three in case of a fourth. Subpicture implication (type [1]) seems to be an attractive candidate for this, as it can be used to define type [4] (if everything is a subpicture of itself) as well as type [2] in the following way:

> $P_1,\ldots, P_n$ [2]-imply Q iff for every R, if P_1 [1]-implies R (that is, if R is a subpicture of P_1), and... P_n [1]-implies R, then Q [1]-implies R.

Unfortunately, it is not possible to define type [3] in terms of type [1] as well, nor is there any other of the four examples which could be used to define all the other ones. We therefore have to use another strategy to argue that the four examples are all special cases of a single pictorial implication relation. Consider the following definition of a pictorial implication relation $\rightharpoonup$:

> $P_1,\ldots, P_n \rightharpoonup Q$ iff the information contained in Q is properly or improperly included in the information in $P_1,\ldots, P_n$.[18]

---

[18] John Corcoran has developed a theory of implications for *propositions* based on the

First of all, note that there are two distinct senses of understanding the informational content of a picture. We might conceive of it as objective information, which is exclusively a feature of the picture, or as subjective information, which depends on the sensory and cognitive apparatus of the perceivers. Two different pictorial tokens might contain the same objective information but different subjective information. An example of this is the case of a photograph and its negative: we might be able to extract more information from one rather than the other, though the objective informational content of the two is exactly the same. For the purposes of this paper, I will restrict myself to objective informational content. The information a picture contains is therefore taken to be identical for all different possible perceivers considered.

Now the informational content of a subpicture thus perceived is obviously properly included in the bigger picture of which it is a part. All the parts of a fresco contain the same information as the fresco they jointly constitute.[19] If we use two pictures to generate an intermediate perspectival view, we lose some information in each of them, so the information present in the intermediate view is properly included in the information contained in the pictures which gave rise to it. Each picture contains the same information as itself. Thus it seems as if this definition in terms of information-content subsumes the four examples of pictorial implication given.

I am aware that my appeal to the information-content of pictures involves a considerable amount of hand-waving. There is at present no satisfactory worked-out theory of the information-content of visual representations (nor, one should hasten to add, is there workable a theory of semantic content on an information-theoretic basis). The

---

very same idea. See his "Information-theoretic Logic," in Concha Martinez, ed., *Truth in Perspective* (Aldershot, UK: Ashgate, 1998), pp. 113–35.

[19] A possible worry one might have with this is the following. Assume two parts of the fresco, *A* and *B*, each show a part of a flower. The information about how many petals the flower has is neither present in *A* nor *B*. Indeed, looking at *A* and *B* we do not even know whether they are parts of the picture of the *same* flower. But this information is present in the entire fresco, consisting of *A* and *B* put togther. So the entire fresco, which is jointly implied by *A* and *B* contains more information than *A* and *B* separately, contrary to the definition, which claims that the information-content in the *implicatum* is always smaller than or equal to that of the *implicans*. The worry can be answered by referring to the distinction between conjunctions and concatenations introduced below. Only the concatenation of *A* and *B* contains the information about how many petals the flower has, but not the conjunction. As only the conjunction, not the concatenation is implied, the problem disappears.

present paper is certainly not the place for coming up with such a new theory. All I am doing here is noting the intuition that if we take a color picture and cut off a bit, or transform it into a black-and-white picture, the resulting picture contains less information than the original, while the results of, say, mirroring it or inversing the colors would not affect the amount of information contained. A theory of the information-content of pictures should be able to account for this intuition, and once such a theory has been developed we can give a clearer account of the background of the definition of pictorial implication in terms of information-content given above. The precise details of the theory do not matter much for our present purposes, as long as there is *any* reasonably systematic way of accounting for the intuitions just given (which seems to me beyond reasonable doubt).

The implication relation between pictures just presented is construed as a mono-categorial implication relation: the signs on either side of the implication sign denote items of a single type or category (namely pictures), in the same way as the objects denoted by the variables flanking $\vdash$ or $\Vdash$ are uniformly formulae or propositions, respectively. However, this is not the only and not even necessarily the most natural way of conceiving of implications between complexes such as pictures or states of affairs. It seems very straightforward to view the implication relation for complexes as poly-categorial. This is because a complex can imply something which is not a complex itself, or can in turn be implied by noncomplexes. A picture's constituents are not pictures in turn, but can become so by adequate supplementation; similarly the state of affairs that John is a bachelor might not just imply the state of affairs that he is male, but also its various constituents (the individual John, the monadic property of being male), which are no states of affairs. Conversely, two nonpictures, two constituents, can be put together to form a picture, thus implying it; two constituents of states of affairs (for example, an individual and a property) can imply the state of affairs resulting from putting the two together.[20]

Although I think that a satisfactory theory of implications dealing with complexes such as pictures or states of affairs should be poly-categorial in the sense just described, I will not pursue this idea further in the present paper. To keep matters simple, I will assume in the following that the symbols on either side of the pictorial implication-sign denote complete pictures, not their constituents.

---

[20] Of course parts of complexes can also imply other parts: constituents of pictures (for example, the complex of an outline together with a shading) can have constituents themselves (the outline, and the shading); the property 'loving Becca' (which is a constituent of a state of affairs) implies a further constituent, the individual Becca.

It is important to note that the implication relation between pictures from a collection G just defined satisfies all of Koslow's six conditions for being an implication relation:

(1) REFLEXIVITY	for all A in G, $A \rightharpoonup A$
(2) PROJECTION	$A_1,\ldots,A_n \rightharpoonup A_k$, for any $k=1,\ldots,n$
(3) SIMPLIFICATION	for all A_i, B in G, if $A_1, A_1, A_2,\ldots,A_n \rightharpoonup B$, then $A_1, A_2,\ldots,A_n \rightharpoonup B$
(4) PERMUTATION	If $A_1, A_2,\ldots,A_n \rightharpoonup B$, then $A_{f(1)}, A_{f(2)},\ldots,A_{f(n)} \rightharpoonup B$ for any permutation f of 1, 2,..., n
(5) DILUTION	If $A_1,\ldots,A_n \rightharpoonup B$, then $A_1,\ldots,A_n, C \rightharpoonup B$ for all A, B, C, in G
(6) CUT	If $A_1,\ldots,A_n \rightharpoonup B$, and $B, B_1,\ldots,B_m \rightharpoonup C$, then $A_1,\ldots,A_n, B_1,\ldots,B_m \rightharpoonup C$ for all A_i, B_j, B, and C.

Every picture implies itself (this follows from the reflexivity of the subpicture relation), and the same holds if we consider any collection of pictures (thus satisfying projection). Repetition of pictures in a collection does not affect their implications, nor does their order. Pictorial implication is monotonic and sequences of implications can be simplified using cut.

At this point, it is instructive to consider another relation between pictures which we might intuitively consider an implication relation of sorts, but which fails to be one, as it only satisfies some of the above conditions.

The individual pictures making up a comic strip which are usually arranged in a narrative sequence might be taken to imply one another. If we consider a sequence of three pictures, the first showing a hammer striking a vase, the second the vase shattering, and the third the cat sleeping next to the vase waking up, it is not implausible to argue that in some way the first picture entails the second, and the second entails the third. The notion of implication in play here is, however, very different from the one studied in logic. While it definitively satisfies reflexivity and projection, and possibly also simplification,[21] permutation fails to be satisfied since the implication in this case also entails a temporal ordering: if we change the sequence of the pictures in the comic strip, the implications will not necessarily be preserved. Dilution also fails: if one picture implies the following, it is not always the case that this implication is preserved when adding another one. The picture of the vase shattering implies that of the cat waking up, but if we interpose a picture between the two which shows that the shattering of the vase was only an image on a television screen in a different room from the cat then the final picture will no longer be implied. Finally, cut is not satisfied. If we "cut out" a sufficiently long sequence between pictures in a comic strip, it is not always the case that the last picture before the cut will imply its new successor, which is the first picture after the cut.

---

[21] This depends on whether we think that a series of repeated pictures in a comic strip can unproblematically be reduced to a single occurrence.

Unlike this "comic-strip implication," the implication relation between pictures based on informational content described above behaves structurally exactly like the implication relation familiar from logic. We will now consider how the different logical operators (conjunction, negation, disjunction, and the hypothetical) between pictures can be defined on the basis of this implication relation.

II.1. Conjunction.

> For any collection of pictures G we define the conjunction $Conj(A, B)$ of two pictures A and B from G to be the weakest picture in G such that $Conj(A, B) \rightharpoonup A$ and $Conj(A, B) \rightharpoonup B$. By saying that it is the weakest picture we mean that if for any other picture X from G, $X \rightharpoonup A$ and $X \rightharpoonup B$ then $X \rightharpoonup Conj(A, B)$.

It is instructive to consider the relation of the *logical* operation of pictorial conjunction with the *spatial* operation of concatenation. Sober regards concatenation as a conjunction operation (*op. cit.*, p. 122), at least when considering typical cases of concatenation.[22] Sober's notion of conjunction is, however, fundamentally different from the one developed here. He employs no framework for defining conjunction directly at the level of pictures, but only via their linguistic descriptions. Sober considers a representation function I such that for every picture p, $I(p)$ is a sentence specifying the information p provides (*op. cit.*, p. 112). It is then argued that an operation + on pictures is the equivalent to the conjunction operation on sentences in case $I(p+q) = I(p) \wedge I(q)$. Neglecting cases of clever blending the operation + can then be identified with concatenation.

---

[22] Concatenation "fails to exactly exemplify pictorial conjunction" (Sober, p. 121) because of the reverse of clever cropping (which we might want to call "clever blending"). If we concatenated two pictures and these would blend into a new picture distinct from the original ones the result would not be a conjunction, as the conjuncts are no constituents of it. See also Robert Howell, "Ordinary Pictures, Mental Representations, and Logical Forms," *Synthese*, XXXIII (1976): 149–74, on pp. 160–61.

Apart from the methodologically unattractive feature of having to switch systems of representations first in order to make sense of logical operations on pictures, Sober's account also faces internal difficulties, as noted by Robert Howell (*op. cit.*, p. 161). If we have two pictures each of which shows two apples next to one another then part of the information each picture provides is that there is exactly one apple to the right of the leftmost apple. Concatenating the pictures so that they show four apples in a row this is no longer the information given, for now there are three apples to the right of the leftmost one.

Of course, Sober could defend himself against this charge by modifying his claim and asserting that pictorial conjunction is not the equivalent of $\wedge$ but of a more complicated truth-functional operation called conjunction′ which only implies some, but not all of its conjuncts′.[23] Whether this defense is successful depends on the properties conjunction′ turns out to have in the end. More problematic seems to be that if we go down that route we are essentially answering a different question: instead of developing a theory of logical operations on pictures, we now inquire which of the claims of its subpictures a picture preserves. This is due to Sober's reluctance to deal with pictures directly and his concentrating instead on their linguistic equivalents.

The main reason, however, why we cannot identify conjunction with concatenation is that conjunction is unique, while concatenation is not. If we have two pictures there are various ways in which we can concatenate them in the plane: putting one above another or below it, one to the left of the other or the other way round, and so on. The difference is not that conjunction is a "purely mental" operation on pictures while concatenation would actually have to move physical pieces around. Both operations deal with pictures, not with paintings, and therefore do not deal with physical objects. To see that the same difference arises in the case of two purely mental operations compare the *mere combination* of some ideas (Paul, Peter, is taller than) and their *combination in a judgment* (Peter is taller than Paul, Paul is taller than Peter): the first operation just specifies that the items in question are put together, the second determines how exactly this is to be done.

Conjunction and concatenation are therefore two distinct cognitive operations on pictures. They can of course be applied in succession: we can form the conjunction of some concatenations, and we can also concatenate conjuncts. This latter is of particular interest when considering logically complex operations on pictures below. Concatenation on its own does not entail any minimality conditions (the sequence of words 'Peter plays well' qualifies as a concatenation of the words 'Peter' and 'plays'), but concatenation *of conjuncts* inherits the minimality condition from the definition of conjunction: the resulting object must be the weakest which can be concatenated from the elements of the conjunction.

---

[23] A straightforward way of accounting for the phenomenon of clever blending would be to say that conjunction′ is nonmonotonic.

While concatenations of pictures are obviously pictures, conjunctions are as well, although (unlike concatenations) their constituents do not stand in a determinate spatial relationship. It makes no sense to ask of the conjunction of the pictures of a cat and a mat whether the first is above the second or the other way round. But this does not keep it from being a picture (although a relatively abstract one) in the same way in which 'a picture of the Annunciation' is a picture, although there is not definite answer to the question regarding which side the angel is coming from.

Conjoining pictures is a cognitive operation which is particularly important when creating paintings. If we are commissioned to paint an Annunciation for a chapel or a cat-mat picture for a philosophy textbook the basis for the design is the weakest picture implying all the necessary subpictures (Mary, the angel, a lily; a cat and a mat), that is, the conjunction of these subpictures. While designing the picture, we then decide how to put these conjunct into a spatial relation with one another; that is, we form a concatenation. On the basis of this we can then proceed to execute the painting.

II.2. Disjunction.

> For any collection of pictures G the disjunction $Disj(A, B)$ of two pictures A and B from G is the weakest picture in G such that for any T in G, if $A \rightarrow T$ and $B \rightarrow T$, then $Disj(A, B) \rightarrow T$.

The disjunction of two pictures is therefore the weakest picture which implies what each of the two disjuncts implies. While the logical operation of conjunction on pictures is central to the creation of paintings, disjunction is central to their comparison, for it gives rise to pictorial abstraction.

Assume we have two identical pictures of a peach, with the difference that the first shows the peach on a table, while the second shows it on a chair. The disjunction of the two is then just the subpicture of the two which shows the peach. We have therefore abstracted the "common element" of the two pictures. Although not every pair of pictures has a disjunction, it is still plausible to regard forming the disjunction as a widespread and natural cognitive operation on pictures. Its precise form depends on the kind of identity conditions we assume for pictures. The strictest conditions would demand a pixel-by-pixel correspondence between pictures, while a more flexible approach might regard two pictures of the same thing as identical, even though they might not have a single pixel in common. On the latter understanding we can see how the notion of the disjunctions of pictures stands behind all attempts to give an account of a particular pictorial motive across a series of pictures. If we, for example, attempt a study of the depiction of windmills in the nineteenth-century English painting, the raw material to consider will be the disjunction of a particular set of paintings, namely those subpictures which are pictures of windmills and which all of a particular collection of paintings have in common.

II.3. Negation.

> For any collection of pictures G the negation *Neg*(*A*) of some picture *A* in G is the weakest picture in G such that for any picture *B* in G, *Neg*(*A*) together with *A* implies *B*.

This definition exploits the fact that something together with its negation implies anything whatsoever. The conjunction of a picture and its negation will be a contradictory picture; a picture of which everything is a subpicture. Whether there are contradictory pictures is a contentious issue,[24] but my account can remain neutral on this point. As collections of objects are not necessarily closed under implication relations there is no need to assume that any such "big picture" will indeed exist.

If our gallery G is rather boring and contains only three pictures of different pieces of fruit: *P*, the picture of a peach, *Q*, that of a quince, and *R*, that of a raspberry then *Neg*(*P*) will just be *Conj*(*Q*, *R*). Once a clear notion of implication between pictures is in place there is therefore nothing inherently mysterious about negative pictures.[25] We do not have to assume that negative pictures depict what is not the case, that they are all crossed by a red diagonal line, or that there is a more subtle hint of negativity present in them. Negative pictures are just pictures which are implicationally related to others in a particular way.

---

[24] See Roy Sorensen, "The Art of the Impossible," in Tamar Gendler and John Hawthorne, eds., *Conceivability and Possibility* (New York: Oxford, 2002), pp. 336–67, for an (indecisive) discussion.

[25] As there is nothing inherently mysterious about negative states of affairs once a mereology has been defined on them. See Meixner, pp. 44–45.

In fact we apply the operation of negation quite naturally in thinking about pictures whenever we consider a particular part in isolation and compare it to the rest of the picture presently at hand. If we discuss a painting which shows some figures in a landscape, and we want to direct somebody's attention to this very landscape we might say: 'Never mind about the men in the foreground—just concentrate on the rest'. What we are asking him to do is to consider the negation of the men in the foreground, that is, the largest subpicture of the painting of which the men in the foreground are no part.

The definition of negation also allows us to address a worry connected with the fact that implication relations are easy to come by: every nonempty set can be equipped with one. This profusion of implication relations is particularly evident when considering the dual of an implication relation. For any implication relation $\rightarrow$ its dual $\rightarrow^*$ is defined as follows: $A_1,\ldots, A_n \rightarrow^* B$ iff for every T, if $A_1 \rightarrow T,\ldots$ and $A_n \rightarrow T$, then $B \rightarrow T$. Now whenever $A \rightarrow B$, $B \rightarrow^* A$.[26] So whenever for two pictures S and T, $S \rightharpoonup T$, it is also the case that $T \rightharpoonup^* S$, where $\rightharpoonup^*$ is defined as

> $P_1,\ldots, P_n \rightharpoonup^* Q$ iff the information contained in Q properly or improperly includes the information in $P_1,\ldots, P_n$.

To make sense of this implication relation consider a particular case of $\rightharpoonup$, namely subpicture (type [1]) implication. This links a picture to all its subpictures. The dual of subpicture implication will do the reverse, and link it to all the pictures of which it is a subpicture. As this "superpicture implication" has the same structural properties as subpicture implications (as specified on page 615) we therefore see that there is some way in which a picture implies not just all the pictures which are parts of it, but also all those of which it is a part. There is no necessary connection between two objects standing in an implication relation and the *implicans* having a greater or equal informational content than the *implicatum*. Similar considerations apply to the other particular cases of $\rightharpoonup$.

Now we might rightly question why the present paper has chosen to study $\rightharpoonup$ rather than $\rightharpoonup^*$, given that the latter is also a perfectly good implication relation. The reason is that $\rightharpoonup$ has a much closer connection with the cognitive operations we employ in thinking about pictures than $\rightharpoonup^*$, and is therefore particularly interesting. This becomes evident in the treatment of negation. Had we chosen $\rightharpoonup^*$ as our pictorial implication relation the roles of conjunction and disjunction would have been reversed: disjunction would have formed the basis of concatenation, conjunction would have been connected with abstraction.[27] This, it seems, would have been little more than a change in terminology. Consider, however, what happens if we replace $\rightharpoonup$ by $\rightharpoonup^*$ in the above definition of negation. Let us once again think of $\rightharpoonup^*$ as superpicture implication. Now instead of the negation of a picture being something such that the conjunction of this negation and the picture together have any other picture as a subpicture, it is now the case that the conjunction of a picture and its negation are a subpicture of any picture whatsoever. It is not easy to make sense of this intuitively, but assume there was an ubiquitous picture U in a collection of pictures, something which was a subpicture of every picture in the collection. Then the negation of a picture A could be conceived of as some sort of thing which hooks onto A and at the same time transforms the compound into U.

---

[26] See Koslow, *A Structuralist Theory of Logic*, p. 62, 9.1.

The important point now is that there is no straightforward conceptual equivalent to negation defined in terms of $\rightharpoonup^*$ we use in thinking about pictures. There is one if we define pictorial negation in terms of $\rightharpoonup$: it is what allows us to isolate particular parts of pictures from others. For this reason $\rightharpoonup$ appears to be a more interesting pictorial implication relation than its dual.

II.4. Hypothetical.

> For any collection of pictures G containing A and B, the hypothetical $Hyp(A, B)$ is the weakest picture in G such that $A, Hyp(A, B) \rightharpoonup B$.

The hypothetical of two pictures is thus that picture which, together with the first one, implies the second. Again, it is easy to find a natural example of hypothetical pictures. Consider the case of reconstructing a picture (a fresco, for example) from an incomplete set of fragments. Let A be a fragment of a fresco and B a photographic reproduction of a detail from the fresco before its partial destruction. Let us also assume that the picture B shows partially overlaps the one A shows. Now if we added a reconstruction of the remaining part to A the resulting complete fresco would imply B. This remaining part therefore fulfils the condition of being the hypothetical of A and B: it is what we have to add to A in order to get B out.

To take another example, consider the picture A of a building from a certain perspective, say, from the north. Suppose we wanted to have a picture B of the same building from the north-west. Various pictures would imply B together with A (a picture of the building from the west, from north-north-west, and so). The weakest of these (if there is one) will be the hypothetical *Hyp*(*A*, *B*).

---

[27] See Koslow, *A Structuralist Theory of Logic*, pp. 115–16.

II.5. Complex Operations. It is now apparent that the application of logical operations to pictures defined in structuralist terms yields natural and familiar cognitive operations we use when thinking about pictures. Conjunctions serves as a basis for concatenation (putting pictures together to form a bigger one), disjunction results in abstraction (extracting the common parts of distinct pictures), negation yields isolation (extracting a particular part of a single picture), and forming hypotheticals gives supplementation (adding the missing parts of a picture).

But these operations are not just employed on their own, they are frequently combined. Restoring a fresco, for example, entails first forming a series of hypotheticals *Hyp*(*A*, *B*), based on a particular extant fragment *A* and some conception of what the original *B* looked like. Subsequently, one concatenates the conjunction of the fragments with their hypotheticals to form the completed picture.

A more pedestrian but equally interesting case is constituted by "spot the difference" pictures frequently found in childrens' magazines. These consist of two nearly identical pictures which differ in a specified number of respects; the goal is to identify all of them. The logical operation required here is that of forming the negation of the disjunction of the two pictures. Taking first all the parts which the pictures have in common, and then identifying the strongest parts which do not imply anything the common parts imply will give us exactly all the subpictures in which the two pictures differ.

Similarly, situations can arise in more serious contexts. In a recent case, it was argued that a landscape painting by the nineteenth-century Russian realist Ivan Shishkin to be auctioned at Sotheby's for £700,000 was really a doctored version of a work by the relatively unknown Dutch Martinus Koekkoek, (valued at about £5000).[28]

The only difference between the two paintings is that the presumed Koekkoek shows some figures where the alleged Shishkin shows none; to compensate, the latter one shows a signature (that of Shishkin) which is lacking in the former. These differences are of course the negation of the disjunction of the two paintings. Different parts of this picture are hypotheticals which can be conjoined with the disjunction of the two paintings ("what they have in common") to turn it either into the Koekkoek or the Shishkin. If the signature was copied from another painting by Shishkin, it constitutes the negation of all the other subpictures in this other painting which was then conjoined with the negation of a small part of the Koekkoek (namely the part where the signature was going to go). It is evident that the two paintings stand in a variety of logical relations which are, however, nothing but a precise articulation of familiar operations like concatenation, abstraction, isolation, and supplementation we use when thinking about pictures.

---

[28] As reported in the *Guardian* (July 10, 2004).

I hope this paper has shown that the notion of implication between pictures, as well as that of logical operations on pictures can be defined in a precise way. Moreover, these logical operations are not just interesting from a purely formal perspective but correspond to cognitive operations which we apply naturally when thinking about pictures.

I expect these observations to have implications for both the theory of representation and the philosophy of mind. Regarding the former a theory of logical relations between pictures will help to overcome the unnecessary fixation of the philosophical study of representation on linguistic representation by showing how logical relations hold between all kinds of representations, whether linguistic or not. Regarding the latter, much of the discussion in the current imagery debate is based on a more or less sharp division between syntactic and image-like forms of representation. Showing the extent to which logical relations hold for both could be the first step towards a unified theory of mental representation which incorporates both the descriptive and the depictive modes as special cases. Both of these topics will, unfortunately, have to be left as subjects of further investigation.

JAN WESTERHOFF

Faculty of Philosophy,
University of Oxford

AGAINST FICTIONAL REALISM

Fictional realists accept an ontology of fictional characters, creatures, places, and times. They are generally led to do so by a form of argument which is familiar from other areas of philosophy. For fictional realists note that there are apparently true sentences which appear to make reference to, or quantify over, fictional objects:

(1) (a) Raskolnikov is a fictional character.
  (b) The character of Raskolnikov was created by Dostoyevsky.
  (c) Raskolnikov is a more realistic character than Alyosha.
  (d) There are fictional characters which could never have been depicted prior to the creation of Raskolnikov.
  (e) There is a fictional character who, for every novel, either appears in that novel or is a model for a character who does.[1]

Now (1a)–(1e) appear to be literally and straightforwardly true. And it seems, *prime facie*, that we should take their syntax at face value. There simply is not an obvious systematic way of paraphrasing (1a)–(1e) that captures their original senses. But in this case, the fictional realist argues, we seem forced to accept that the occurrences of the name 'Raskolnikov' in (1a)–(1d) refer to a fictional object and that the quantifiers in (1d)–(1e) range over fictional objects. And this commits us to the existence of fictional objects.

The suggestion that the truth of (1d)–(1e) entails the existence of fictional objects is reinforced by two further observations. First,

---

[1] This sort of argument can be found in, for example, Charles Crittenden, *Unreality: The Metaphysics of Fictional Objects* (Ithaca, NY: Cornell, 1991), p. 95; Robert Howell, "Fiction Objects," *Poetics*, VIII (1979): 129–77, see p. 152; Peter Lamarque and Stein Haugom Olsen, *Truth, Fiction, and Literature* (New York: Oxford, 1994), pp. 96–99; Terence Parsons, *Nonexistent Objects* (New Haven: Yale, 1980), pp. 52–54; Amie Thomasson, *Fiction and Metaphysics* (New York: Cambridge, 1998), pp. 5–6; and Peter van Inwagen, "Quantification and Fictional Discourse," in Anthony Everett and Thomas Hofweber, eds., *Empty Names, Fiction, and the Puzzles of Non-Existence* (Stanford: CSLI, 2000), pp. 235–47, see pp. 243–44, as well as "Creatures of Fiction," in his *Ontology, Identity, and Modality* (New York: Cambridge, 2001), pp. 37–56, on p. 43, and "Existence and Ontological Commitment," in Michael Loux and Dean Zimmerman, eds., *Oxford Handbook of Metaphysics* (New York: Oxford, 2003), pp. 131–57, see pp. 136–38. Strict Quineans will insist that our ontological commitments are revealed, not by the natural language sentences we hold true, but rather by the way we choose to paraphrase these into First Order Logic but I will ignore this complication here.

0022-362X/05/0212/624–49

(1d)–(1e) can participate in certain apparently valid inferences, such as the inference from (1e) to the sentence:

(2) If no character appears in every novel then some character is modeled on another character.

The validity of this inference, it is suggested, requires us to understand the quantification involved as objectual quantification over fictional objects. Consequently, in so far as we take (1e) to be true, we seem committed to there being true sentences involving objectual quantification over fictional objects and hence to there being fictional objects.[2] Second, each of (1a)–(1e) appears to entail directly the sentence:

(3) There are fictional objects.[3]

In so far as we take (1a)–(1e) to be true we therefore seem committed to the truth of (3) and hence to the existence of fictional objects.

Let us call sentences such as (1a)–(1e) which appear to make reference to, or quantify over, fictional objects *Fictional Object Sentences* or *FOS* for short. Now fictional realists sometimes also offer analogous arguments for their position that invoke other sorts of linguistic data. Thus, for example, fictional realists sometimes argue that we are committed to fictional realism by the apparent truth of object-fictional sentences such as (4a) which appear to say what is true in the world of the relevant story, fictional intensional transitives such as (4b), and sentences comparing fictional and real objects such as (4c):

(4) (a) Raskolnikov is a student.
(b) Putin is searching for Raskolnikov.
(c) Putin is less neurotic than Raskolnikov.[4]

However, these cases are far more controversial. It is plausible that utterances of object-fictional sentences do not involve us actually asserting anything but rather involve our merely pretending to make assertions or at least our making assertions which describe the content of a pretense rather than the nature of the real world.[5] And if this is true it is not immediately clear how utterances of sentences such as (4a) *could* commit us to anything. Moreover, alternative treatments of the relevant comparatives and intensional transitives are available that do not seem to involve a commitment to fictional objects.[6] Of course, it might ultimately turn out that the only way to make sense of (4a)–(4c) requires our taking them to make reference to fictional objects. But this is not at all obvious and requires further argument. In what follows, then, I will confine my attention to the argument for fictional realism that is based upon the apparent truth of *FOS*.

---

[2] See van Inwagen, "Quantification and Fictional Discourse," p. 244.

[3] See van Inwagen, "Existence and Ontological Commitment," p. 137.

[4] See, for example, Lamarque and Olsen, *Truth, Fiction, and Literature*, pp. 97–101; Karel Lambert, *Meinong and the Principle of Independence* (New York: Cambridge, 1983), p. 144; and Parsons, *Nonexistent Objects*, pp. 32–38.

[5] The classic statement of the view that utterances of object-fictional sentences should be understood as pretend assertions can be found in John Searle, "The Logical Status of Fictional Discourse," in his *Expression and Meaning* (New York: Cambridge, 1979), pp. 58–75; see Lamarque and Olsen, *Truth, Fiction, and Literature*, pp. 63–69, for a

Now there is some disagreement among fictional realists as to how we should understand the nature of fictional characters. Some, such as David Braun, Nathan Salmon, and Amie Thomasson, take fictional objects to be abstract human artifacts created by authors.[7] Nicholas Wolterstorff and Ed Zalta take fictional objects to be eternal abstract Platonic objects.[8] And Peter van Inwagen takes fictional characters to be abstract objects but does not commit himself as to whether they are created or eternal.[9] Moreover, there are those, such as Alexius Meinong, Terence Parsons, and Richard Routley, who take fictional objects to be nonexisting concrete objects.[10] We can ignore these differences for the moment, however, since the objections I will raise below apply equally to all of these approaches.

---

discussion of Searle's position. See Stuart Brock, "Fictionalism about Fictional Characters," *Noûs*, XXXVI, 1 (2002): 1–21; van Inwagen, "Existence and Ontological Commitment," pp. 131–35; and Kendall Walton, *Mimesis as Make-Believe* (Cambridge: Harvard, 1990), pp. 396–405, for discussions of the ontological implications, or lack thereof, of object-fictional sentences.

[6] For examples of the former, see Mark Crimmins, "Hesperus and Phosphorus: Sense, Pretense, and Reference," *Philosophical Review*, CVII, 1 (1998): 2–48; and Walton, *Mimesis as Make-Believe*, p. 414. For examples of the latter, see Mark Richard, "Commitment," in James Tomberlin, ed., *Philosophical Perspectives 12* (Cambridge: Blackwell, 1998), pp. 255–81, and "Seeking a Centaur, Adoring Adonis: Intensional Transitives and Empty Terms," in Howard Wettstein and Peter French, eds., *Figurative Language: Midwest Studies in Philosophy, Volume XV* (Malden, MA: Blackwell, 2001), pp. 103–27.

[7] See Braun, "Empty Names, Fictional Names, Mythical Names," *Noûs*, XXXIX, 4 (2005): 596–631; Salmon, "Nonexistence," *Noûs*, XXXII, 3 (1998): 277–319; and Thomasson, *Fiction and Metaphysics*, "Fictional Characters and Literary Practices," *British Journal of Aesthetics*, XLIII, 2 (2003): 138–57, and "Speaking of Fictional Characters," *Dialectica*, LVII, 2 (2003): 205–23.

[8] See Wolterstorff, *Works and Worlds of Art* (New York: Oxford, 1980); and Zalta, *Abstract Objects* (Boston: Reidel, 1983), and *Intensional Logic and the Metaphysics of Intensionality* (Cambridge: MIT, 1988).

[9] See van Inwagen, "Quantification and Fictional Discourse," "Creatures of Fiction," and "Existence and Ontological Commitment."

[10] See Meinong, "Über Gegenstandstheorie," in his *Untersuchungen zur Gegenstandstheorie und Psychologie* (Leipzig: Barth, 1904), and *Über die Stellung der Gegenstandstheorie im System der Wissenschaften* (Leipzig: Voigtlander, 1907); Parsons, "A

Of course, if we accept fictional characters into our ontology then we face the task of determining precisely which fictional characters there are. Now it seems obvious that which fictional characters we take to occur in a story will depend upon what the world of that story is like. And, in fact, I suggest that the following principles have a near platitudinous status:

(P1) If the world of a story concerns a creature *a*, and if *a* is not a real thing, then *a* is a fictional character.

(P2) If a story concerns *a* and *b*, and if *a* and *b* are not real things, then *a* and *b* are identical in the world of the story iff the fictional character of *a* is identical to the fictional character of *b*.

Some of our intuitions about fictional characters may, arguably, be potentially revisable. But (P1) and (P2) seem so fundamental to our conception of a fictional character I doubt very much that any account which rejected (P1) or (P2) would deserve to be counted as an account of fictional characters but should rather be regarded as an account of some other sort of entity. Note in any case that our intuitions that (P1) and (P2) are true seem at least as strong as our intuitions that (1a)–(1e) are true. Moreover, our acceptance of many *FOS* appears to result from our applying (P1) and/or (P2). Thus, for example, I suspect that our intuitions that (P1) is true are at least as strong as our intuitions that (1a) is true and that we only judge (1a) to be true because we accept (P1) and take the world of *Crime and Punishment* to contain Raskolnikov. I would argue, then, that the fictional realist cannot reject (P1) or (P2) without thereby undermining our motivation for accepting fictional realism in the first place. In what follows I shall therefore assume that the fictional realist is committed to (P1) and (P2).

Of course, as they stand, these principles are not terribly helpful since they presuppose an account of what it is for something to be the case in the world of a story. And this is a complex matter depending not merely upon what the relevant text says but upon the interpretation we bring to bear upon that text. Still, let us put these difficulties aside for the moment. In what follows, I shall argue that a commitment to (P1) and (P2) is enough to get the fictional realist into difficulties. In particular, I shall argue, authors may leave certain

---

Prolegomenon to a Meinongian Semantics," this JOURNAL, LXXI, 16 (September 19, 1974): 561–80, *Nonexistent Objects*, and "Are There Nonexistent Objects?" *American Catholic Philosophical Quarterly*, XIX (1982): 365–71; and Routley, *Exploring Meinong's Jungle and Beyond: Department Monograph 3, Philosophy Department* (Canberra: Australian National University, 1980).

things unspecified about the world of their story including whether certain creatures count as identical or distinct in that world and which creatures exist in that world. Given (P1) and (P2), this sort of underspecification within a story gives rise to ontic indeterminacy concerning which fictional characters occur within that story. Moreover, I shall argue, if the laws of logic and identity fail in the world of a story, these failures may infect the fictional characters occurring in that story. In short, given (P1) and (P2), the fictional realist seems committed to certain pernicious forms of indeterminacy and to objects that flout the laws of logic and identity. These are obviously undesirable commitments. And, since these consequences do not arise for a pretense-theoretic account of fictional objects, I shall argue that we should favor a pretense-theoretic account over fictional realism. I conclude by briefly comparing the objections I raise in this paper to those raised by Bertrand Russell to Meinong.

I. OBJECTION 1: ONTIC INDETERMINACY

Let us turn to our first objection to the fictional realist. Now it is widely recognized that if there are fictional objects then it will at least sometimes be a vague or indeterminate matter as to whether two fictional objects are identical.[11] This is particularly clear when we consider questions of intertextual identity. Consider, for example, the question of whether Marlowe's Faust is the same character as Goethe's Faust. Now I suspect our concept of fictional characters is insufficiently precise to provide an answer to this question and we have no principled pre-theoretic way of deciding it one way or the other. It seems plausible, then, to regard it as *indeterminate* whether these Faust characters are identical or distinct.

Many fictional realists seem willing to accept such cases of intertextual indeterminacy and, as they stand, I doubt these cases pose much of a problem to the fictional realist. For it is common to distinguish between benign cases of indeterminacy which arise as the result of imprecision in our concepts and language and pernicious cases of ontic indeterminacy where the nature of the world is itself indeterminate. And I think the fictional realist can plausibly argue that the cases of indeterminacy noted above result because of the imprecision in our concept of a fictional character. Our concept of a fictional character is simply not sufficiently precise to settle whether we have two Faust characters or one but, were we so inclined, we could precisify it so as to settle this question.

---

[11] See, for example, Lamarque, "Objects of Interpretation," *Metaphilosophy*, XXXI, 1/2 (2000): 96–124 (see in particular p. 120); and Thomasson, "Fictional Characters and Literary Practices," p. 155.

Nevertheless, a little reflection suggests that the fictional realist is also committed to cases of genuine ontic indeterminacy. For a story might describe two characters in such a way that it is left indeterminate whether or not they are identical:

> *Frackworld*: No one was absolutely sure whether Frick and Frack were really the same person or not. Some said that they were definitely two different people. True, they looked very much alike, but they had been seen in different places at the same time. Others claimed that such cases were merely an elaborate hoax and that Frick had been seen changing his clothes and wig to, as it were, become Frack. All that I can say for certain is that there were some very odd similarities between Frick and Frack but also some striking differences.

I think it is pretty clear that in this story it is left indeterminate as to whether Frick is Frack. But in this case (P2) entails that it is indeterminate whether the Frick-character is identical to the Frack-character.[12] Note that this indeterminacy does not arise because our concept of a fictional character is vague and imprecise. No attempt to refine or precisify our concept of a fictional character can remove the indeterminacy in Frackworld and, granted (P2), this indeterminacy in Frackworld will carry over into an indeterminacy concerning whether or not the Frick-character is the same as the Frack-character.

This ontic indeterminacy is potentially problematic for the fictional realist for there is a well-known argument from Gareth Evans which appears to show that it cannot be an indeterminate matter as to whether a is b.[13] The argument runs as follows. Suppose that it was indeterminate as to whether a was b. Then b would have the property of being indeterminately identical to a. But, since a is determinately identical to a, a does not have the property of being indeterminately identical to a. So there is a property which b has but a lacks. So by Leibniz's Law $a \neq b$. Now, as it stands, this is not a strict *reductio* of our assumption. But it is surely an undesirable consequence for those who wish to maintain that it is indeterminate as to whether a is b. Moreover, we may transform Evans's argument into a *reductio* of the assumption that it is definitely the case that it is indeterminate whether *a* is *b* as follows. Granted that Evans's original argument preserves definiteness, it follows from the assumption that it is definitely the case that it is indeterminate whether *a* is *b* that *a* is definitely distinct from *b*. But our assumption also entails that it is indeterminate whether *a* is *b* which, plausibly, entails that *a* is *not* definitely distinct from *b*.

---

[12] This is granted, of course, the plausible assumption that the biconditional in (P2) preserves indeterminacy.

[13] Evans's original argument is in his "Can There Be Vague Objects?" *Analysis*, XXXVIII (1978): 208. There is an enormous body of literature on Evans's argument to which I cannot hope to do justice here, but see Timothy Williamson, "Vagueness in Reality," in Loux and Zimmerman, eds., for an excellent discussion of the relevant issues and references to the relevant literature.

Of course, it might turn out in the end that there is something wrong with Evans's argument. But it is worth heading off at least one potential response on the part of the fictional realist here. Some have thought that Evans's argument may be blocked by a so-called fuzzy logic invoking degrees of truth and the fictional realist might hope to avail herself of this response.[14] The problem here is that it is completely unclear how we could assign any degree of truth at all to the claim that Frick is Frack in our story. We cannot say whether that claim is very true, or quite true, or quite false, or very false, or somewhere in between. We cannot even say if any of these options are truer than any of the others. Fuzzy logic may work well for cases of vagueness where we have a series of borderline cases possessing a property to a greater or lesser degree. It does not work terribly well with the sort of penumbra-free indeterminacy generated by *Frackworld*. Anyway, I think it is clear that those who wish to reject Evans's argument must argue their point. And until the fictional realist discharges this burden so much the worse for her position.

Fictional realists also seem to be committed to a rather different form of ontic indeterminacy. For they appear committed, not merely to cases of vague identity, but also to cases of vague existence or indeterminate being. Consider Tatyana Tolstaya's recent novel *The Slynx*. Now in the end, I think, it is pretty much left open whether or not there really is a Slynx in Tolstaya's novel. It seems to be an indeterminate matter as to whether the world of that novel contains a Slynx. But then, given (P1), it follows that it is an indeterminate matter as to whether there is such a fictional creature as the Slynx.[15]

The problem here is that it is not clear how the being of the Slynx-creature could be an indeterminate matter. Three possibilities suggest themselves. The fictional character of the Slynx might be a special sort of object with an indeterminate ontological status so it does not definitely have being but does not definitely lack being either. It might be indeterminate as to whether the property of being the Slynx-character is instantiated. Or perhaps it is indeterminate as to whether the state of affairs of there being a Slynx-character obtains. I do not see what else it could be for it to be indeterminate as to whether there is a Slynx-character. But in the present case none of these will do.

---

[14] For an application of fuzzy logic to Evans's argument, see B. Jack Copeland's excellent "Fuzzy Logic and Vague Identity," this JOURNAL, XCIV, 10 (October 1997): 514–34. For classic expositions of fuzzy logic, see Joseph Goguen, "The Logic of Inexact Concepts," *Synthese*, XIX (1969): 325–73; Kenton Machina, "Vague Predicates," *American Philosophical Quarterly*, IX (1972): 225–33, and "Truth, Belief, and Vagueness," *Journal of Philosophical Logic*, V (1976): 47–78; and Lofti Zadeh, "Fuzzy Sets," *Information and Control*, VIII (1965): 338–53.

[15] I assume, of course, that the biconditional in (P1) preserves indeterminacy.

The first option fails because it is not clear that those who postulate objects of indeterminate ontological status can even coherently formulate their position. In order to do this they must, presumably, maintain that there are objects which do not determinately have being but which do not determinately lack being either. But, of course, in so far as they commit themselves to *there being* such objects, they appear to commit themselves to those objects having a determinate ontological status after all.[16]

It might be replied, against this objection, that those who postulate objects of indeterminate ontological status should not maintain that there *are* such objects but rather merely that it is indeterminate as to whether there are any such objects. This might be so. But in the present context this reply simply pushes our problem one stage back. We wanted an account of what it could be for it to be indeterminate as to whether there is a Slynx-character. The suggestion we were considering was that the Slynx-character might be an object with an indeterminate ontological status. In other words we invoked the idea that the Slynx-character might be an object with an indeterminate ontological status in order to make sense of the claim that it is indeterminate whether the Slynx-character exists.

I argued, however, that we cannot make sense of there *being* a Slynx-character of indeterminate ontological status. Rather we must hold that it is indeterminate whether there is a Slynx-character of indeterminate ontological status. But this means that we still need an account of what it is for it to be indeterminate as to whether there is a Slynx-character. And this is the very phenomenon that the postulation of ontologically indeterminate objects was supposed to explain. We have gone around in a circle and the postulation of ontologically indeterminate objects does no explanatory work here. I conclude that the fictional realist cannot provide an account of what it is for it to be indeterminate whether there is a Slynx-character by maintaining that the Slynx-character is an object of indeterminate ontological status.

---

[16] Van Inwagen makes a similar point in his *Material Beings* (Ithaca, NY: Cornell, 1990), p. 240.

The second option fails because, in order for it to be indeterminate as to whether a property is instantiated, there must surely be an object or set of objects such that it is indeterminate as to whether they instantiate that property. At any rate, if every object determinately fails to instantiate a given property it is not clear how it could be indeterminate whether that property is instantiated. Now if the property of being the Slynx-character is indeterminately instantiated then it must be indeterminately instantiated by a fictional character in Tolstaya's novel. But every fictional character occurring in Tolstaya's novel which has a determinate ontological status determinately does not instantiate the property of being the Slynx-character. So the property of being the Slynx-character is not indeterminately instantiated by any objects that have a determinate ontological status. If it is indeterminately instantiated at all, then, it must be indeterminately instantiated by an object which does not have a determinate ontological status. So, if the suggestion we are considering is to work, we must accept that there are objects which lack a determinate ontological status. But, of course, we have just rehearsed the problems facing the view that there are such things.

The third option fails because, if it is indeterminate as to whether the state of affairs of there being a Slynx-character obtains, it must surely be indeterminate as to whether the property of being the Slynx-character is instantiated. After all, if that property was determinately instantiated or determinately failed to be instantiated then it is hard to see how it could be an indeterminate matter as to whether the state of affairs of there being a Slynx-character obtained. But we have already seen that, in the present instance, we cannot make sense of it being indeterminate as to whether the property of being the Slynx-character is instantiated.

Of course, there might be some other way of understanding how it could be indeterminate as to whether there is a Slynx-character that we have not considered. But, if so, I do not see what it is. And the fictional realist is clearly faced with the task of explaining it. Until the fictional realist discharges this burden, so much the worse for her position.

The two arguments just given assume (P2) and (P1) respectively and, although I argued earlier that the fictional realist must accept these principles, a fictional realist might nevertheless hope to block my arguments by rejecting them. However, it is worth pointing out that this strategy faces a problem. For, even if we block my arguments by rejecting (P1) and (P2), we still face the problem of deciding whether the Frick-character is the same as the Frack-character and whether there is a Slynx-character. And there seems no principled way in which we might decide these questions. In each case we have no more reason to choose one of the options than the other and whatever choice we make will be unacceptably *ad hoc*. Hence, I would argue, simply rejecting (P1) and (P2) will not save the fictional realist.

Before concluding our discussion of indeterminacy, it is worth emphasizing that the cases we have been discussing really generate two problems for the fictional realist. First, of course, the fictional realist seems committed to two problematic varieties of ontic indeterminacy. But even those who are willing to embrace ontic indeterminacy face a second and perhaps deeper problem here. For, if fictional realism were true, ontic indeterminacy would be uncomfortably easy to come by. To bring this point out imagine, for a moment, that God created the world so that it was completely precise and determinate, so that there was no ontic indeterminacy of any form. If fictional realism was true then human beings could still generate cases of ontic indeterminacy simply by writing fiction. This seems disquieting. Surely we do not have this degree of control over the metaphysical nature of the world. And we should surely be wary of any view which entails that we do.

II. OBJECTION 2: LOGICAL INCOHERENCE

Let us turn to our second objection to the fictional realist. The difficulty here is that a story might describe an impossible world in which the laws of logic or identity fail. But since, by (P1) and (P2), what exists in the world of a story determines which fictional characters occur in that story, various impossibilities within the world of a story may infect the fictional characters that occur in that story. Here are two examples.[17] In the first story the law of noncontradiction fails. In the second, the symmetry of identity fails:

> *Dialethialand*: When she arrived in Dialethialand, Jane met Jules and Jim. This confused Jane since Jules and Jim both were, and were not, distinct people. And this made it hard to know how to interact with them. For example, since Jules both was and was not Jim, if Jim came to tea Jules

---

[17] These examples might be multiplied. With sufficient ingenuity we might construct stories where, for example, the law of excluded middle, the reflexivity of identity, the transitivity of identity, and the laws of arithmetic, fail to apply to identity statements involving the denizens of the stories and hence fail to apply to the fictional characters occurring in those stories. Indeed, Tamar Szabo Gendler's *The Tower of Goldbach* (see her "The Puzzle of Imaginative Resistance," this JOURNAL, XCVII, 2 (February 2000): 55–81) provides a case where, arguably, 7 fictional characters plus 5 fictional characters both does and does not equal 12 fictional characters.

> both would and wouldn't come too. This made it hard for Jane to determine how many biscuits to serve. Then Jane realized what to do. She needed both to buy and not to buy extra biscuits whenever Jim came. After that everything was better.
>
> *Asymmetryville*: As soon as he got up in the morning Cicero knew that something was wrong. It was not that he was distinct from Tully. On the contrary, just as always he was identical to Tully. It was rather that while he was identical to Tully, Tully was distinct from him. In other words, some time during the night (he could not tell exactly when) the symmetry of identity failed. This had some rather annoying consequences. When Cicero got paid Tully could spend the money but not vice versa. Tully got fat off the food Cicero ate and gave up dining himself. And Tully was praised for Cicero's denunciation of Catiline although he himself had slept through the whole affair. It was enough to test Cicero's Stoicism to the limits. Then something happened that changed everything. Cicero's political enemies who knew that Cicero was Tully mistook Tully for Cicero and murdered him. At first it seemed as if Tully had died. But then Cicero realized that since he was alive and he was Tully, Tully was alive too. Tully was understandably grateful and reformed his ways. After that Cicero and Tully lived together happily.

These magical realist stories describe worlds that are supposed to flout the laws of logic. In the world of *Dialethialand*, Jules both is and is not Jim and in the world of *Asymmetryville*, Cicero is Tully though Tully is not Cicero. The question arises, then, as to which fictional characters occur in these stories.

If, as I suggested, the fictional realist is committed to (P1) and (P2) then there seem only two responses open to her if she wants to avoid a commitment to objects that flout the laws of logic and identity. She could maintain that my stories do not contain any fictional characters at all. Or she could allow that they contain fictional characters but maintain that these characters do not after all flout the laws of logic and identity. Let us call the first response the *No Character* response and the second response the *Coherent Character* response. I will consider and reject each response in turn.

Given (P1) and (P2), the No Character response requires the fictional realist to deny that *Dialethialand* and *Asymmetryville* describe fictional worlds that contain people. Now in so far as *Dialethialand* and *Asymmetryville* describe fictional worlds those worlds certainly appear to contain people. So I think that if the fictional realist is to maintain the No Character response she had better argue that, despite appearances to the contrary, *Dialethialand* and *Asymmetryville* do not succeed in describing fictional worlds in the first place.

Unfortunately, this claim seems highly implausible and it is not clear how it might be defended. Of course, the fictional realist could simply insist that fictional worlds must obey the laws of logic and identity. But without some independent motivation this seems a terribly *ad hoc* maneuver and I doubt it could be maintained. After all, many stories contain lurking or explicit contradictions but we nevertheless take them to provide perfectly good descriptions of fictional worlds.

A more interesting variant of this response on the part of the fictional realist would hold that we cannot really imagine the worlds which *Dialethialand* and *Asymmetryville* attempt to describe because this would require us to imagine something logically incoherent and then argue that a story counts as successfully describing a fictional world only in so far as we can imagine that world.[18] But this response also faces its problems. While *Dialethialand* and *Asymmetryville* are no doubt strange stories, I think many readers are able to engage imaginatively with them. After all, I suspect that many readers will find Jane's response to the biscuit problem very appropriate, given that she is in Dialethialand. And I suspect that many will also find the way Cicero and Tully resolve their conflict appropriate. But the very fact that we find such things appropriate is surely a sign that we are able to engage with my stories and have succeeded, at least to some extent, in imagining the worlds they portray.

Of course, it might be objected, we cannot really imagine the worlds of my stories because we cannot really imagine in detail what the world would have to be like in order for my stories to be true. Rather, it seems, when we try to imagine the worlds described by *Dialethialand* and *Asymmetryville*, we will only focus upon certain aspects of those worlds at any one time. At different times we may focus upon aspects of these worlds that are not logically compatible with each other. But, at any one time, what we imagine is logically coherent. And we are unable to bring these different islands of coherence together and imagine what Dialethialand and Asymmetryville are really like in any detail or completeness.

---

[18] We might, for example, take my stories to be describing conceptual impossibilities and agree with Kathleen Stock that we cannot imagine conceptual impossibilities; see her "The Tower of Goldbach and Other Impossible Tales," in Matthew Kieran and Dominic Lopes, eds., *Imagination, Philosophy, and the Arts* (New York: Routledge, 2003), pp. 107–24. Of course, Stock's arguments raise the question of what counts as a conceptual impossibility. But, putting this issue to one side, I think there are serious reasons to doubt their cogency. Brian Weatherson, "Morality, Fiction, and Possibility," *The Philosopher's Imprint*, IV (2004): 1–27 (see especially pp. 10–11) raises some well-taken worries about some of Stock's arguments and I give some further reasons to suppose can imagine the logically incoherent below. See Szabo Gendler's "The Puzzle of Imaginative Resistance," and "On the Relationship between Pretense and Belief," in Kieran and Lopes, pp. 139–40, for a defense of our ability to imagine conceptual impossibilities. Weatherson, "Morality, Fiction, and Possibility," argues that we are able to imagine certain impossibilities although he is more cautious than Gendler about our ability to imagine *conceptual* impossibilities.

I grant that we cannot imagine in any detail or completeness what Dialethialand or Asymmetryville would be like. But this seems equally true of many other stories and does not appear to prevent our imagining the worlds of those stories. When we watch Aristophanes we have no difficulty imagining that the birds build a Cloud-Cuckoo-Land between heaven and earth. Generations of readers and viewers have had no difficulty imagining that Dr. Who's Tardis is a small police phone box on the outside but contains a vast space within. And generations of children have had no difficulty imagining that Winnie the Pooh and Piglet are talking animals. But when we stop and try to imagine in detail how the world would have to be if any of these things were the case then our imaginations are likely to balk. A great deal of the world's fiction will generate imaginative resistance if we try to delve too closely into what the world would have to be like in order for the relevant stories to be true but this does not prevent us from imaginatively engaging with that fiction and imagining the word it describes. So I grant that we cannot imagine in detail what the worlds portrayed in *Dialethialand* and *Asymmetryville* would be like but I do not think this precludes our imaginatively engaging with those stories or our imagining those worlds to be real.[19]

There is obviously a great deal more to say here but for the moment I want to point out two reasons to suppose that we are, in general, able to imagine logically incoherent or inconsistent states of affairs. First, observe that we do seem able to pretend or imagine that some logically impossible state of affairs holds in order to see what follows from them or in the course of a *reductio* proof that no such state of affairs can obtain. I might, for example, imagine that Cicero is Tully but Tully is distinct from Cicero in the course of a *reductio* proof aimed at establishing the impossibility of this. But if I can imagine this when constructing a *reductio* then surely I can also imagine it when reading fiction. Second, observe that some humans do genuinely seem able to *believe* things that are logically incoherent or inconsistent. Dialetheists believe that certain contradictions are both true and false. Meinongians believe in nonexistent objects. And some of us believe that The

---

[19] Szabo Gendler makes a closely related point in her "The Puzzle of Imaginative Resistance," pp. 69–70.

Trinity is simultaneously both one and three. But if humans can *believe* logically incoherent or inconsistent things then surely we can imagine or pretend those things. If we can believe that The Trinity is both one and three then we can surely pretend that Jules both is and is not Jim.

So much the worse, then, for the No Character response. But what about the Coherent Character response? As I see it there are two problems with this response. We can bring these out by considering how the response applies to *Dialethialand* but analogous points may be made with respect to *Asymmetryville*. The first problem is we have to decide whether we are going to take Jules to be identical to Jim or whether we are going to take Jules to be distinct from Jim. And there seems no possible reason to favor one of these options over the other. It seems *ad hoc* and unmotivated to maintain that Jules and Jim are identical and reject the claim that Jules and Jim are distinct. But it seems equally *ad hoc* and unmotivated to maintain that Jules and Jim are distinct and reject the claim that Jules and Jim are identical. The second problem here is that neither of these options really does justice to my story. In so far as we deny that Jules and Jim are distinct or deny that they are identical we seem to get something badly wrong about the world of my story. Neither of these options, for example, allows us to make good sense of Jane's Dialetheist solution to the biscuit problem. In so far as we find Jane's solution very appropriate in the context of my story, I think we must imagine that Jules both is and is not Jim. The only way to do justice to my story then, I would suggest, is to accept that in the world of *Dialethialand*, Jules both is and is not Jim.

I conclude that neither the No Character response nor the Coherent Character response work. We must accept that, in the world of *Dialethialand*, Jules both is, and is not, identical to Jim. This, together with (P1) and (P2), entails that the fictional object corresponding to Jules both is and is not identical to that corresponding to Jim. Moreover we must accept that, in the world of *Asymmetryville*, Cicero is Tully but Tully is distinct from Cicero. This, together with (P1) and (P2), entails that the fictional object corresponding to Cicero is identical to the fictional object corresponding to Tully but that the latter object is distinct from the former. The fictional realist, then, appears to be committed to the existence of logically incoherent objects.

Once again there are two reasons to be unhappy about this outcome. A commitment to logically incoherent objects is in itself obviously extremely problematic. But there is a second problem facing the fictional realist here. For if fictional realism were true then true inconsistencies and violations of the laws of identity would be uncomfortably easy to come by. If God created a world in which the law of noncontradiction and the laws of identity otherwise held, we would nevertheless be able to violate these laws simply by making up stories like *Dialethialand* and *Asymmetryville.* Surely we do not have this degree of control over the laws of logic and identity. And surely we should be wary of any view which allows that we do.

III. PRETENSE THEORY

In the light of these problems I think that we would be well advised to abandon fictional realism and exclude fictional objects from our ontology. But in this case we need an alternative account of how we should understand (1a)–(1e) and our talk and thought about fictional objects in general. Moreover, of course, our account should allow us to avoid the problems facing fictional realism. In what follows I will sketch a view that I find plausible and that meets these desiderata.

I am going to argue that we should, in a certain sense, take a fictionalist view of fictional characters so I want to begin by briefly considering how we should understand the nature of fiction. I am going to assume that our engagement with fiction should be understood in terms of our engaging in some sort of imaginative act, pretense, or game of make-believe.[20] More precisely, I will assume that our engagement with a fictional text involves our pretending that the world is as that fictional text portrays it. But I do not want to commit myself to the details of any more specific view beyond this for the moment nor do I want to commit myself to any particular account of the imagination, pretense, or make-believe.[21] The reader should feel free to substitute whatever variant of my account she prefers and to adopt her favored account of imagination, pretense, and make-believe. Anyway, the basic idea here is that just as little Johnny might pretend that his bicycle is a horse, and just as little Sally might pretend that she is a Native American telling the story of her tribe, so, when we read *Emma*, at least in the simplest case, we pretend that we are reading a factual narrative and we imagine that what we read really took place.[22] Utterances of object-fictional sentences such as (4a) are to be understood as being made within the scope of these sorts of pretense. So in uttering (4a) we are not making an assertion about the real world. Rather we are pretending that the world is as it is portrayed in *Crime and Punishment* and we utter (4a) in order to describe it.

---

[20] This sort of view is defended by, amongst others, Greg Currie in his *The Nature of Fiction* (New York: Cambridge, 1990), and Walton in *Mimesis as Make-Believe.*

[21] No doubt there is a lot of work still to be done in unpacking these concepts but I think that we have a solid pre-theoretic grasp upon these notions and there is an increasing body of interesting work in philosophy and psychology on pretense and the imagination. See, for example, Currie, "Imagination and Simulation," in Martin Davies and Tony Stone, eds., *Mental Simulation* (Cambridge: Blackwell, 1995), pp. 151–69, "The Paradox of Caring," in Mette Hjort and Sue Laver, eds., *Emotion and the Arts* (New York: Oxford, 1997), pp. 63–77, and "Pretense, Pretending, and Metarepresenting," *Mind and Language,* XIII, 1 (1998): 35–55; Currie and Ian Ravenscroft, *Recreative Minds: Imagination in Philosophy and Psychology* (New York: Oxford, 2003); Szabo Gendler, "On the Relationship Between Pretense and Belief"; Paul Harris, "Imagining and Pretending," in Davies and Stone, eds., and *The Work of the Imagination* (Malden, MA: Blackwell, 2000); Alan Leslie, "Pretense and Representation," *Psychological Review,* XCIV

Next I want to note a phenomenon which suggests that our talk and thought about fictional characters, the sort of talk and thought that fictional realists take to be about real fictional objects, is a natural extension of this sort of imaginative engagement with fiction. For observe that critics are given to saying such things as:

(5) The most famous of Conan Doyle's creations is the great detective Sherlock Holmes who lives in Baker Street.

Now (5) simultaneously talks about Holmes both as if he were a human detective living in Baker Street and as if he were a fictional object created by Conan Doyle. In other words, (5) simultaneously talks about Holmes from the perspective of the reader who is caught up in the world of the Holmes stories and from the perspective of the critic who is engaged in literary criticism. In fact, a little reflection

---

(1987): 412–26; Angeline Lillard, "Making Sense of Pretence," in Charlie Lewis and Peter Mitchell, eds., *Children's Early Understanding of Mind: Origins and Development* (Hillside, NJ: Lawrence Erlbaum, 1994), pp. 211–34, "Pretending, Understanding Pretense, and Understanding Minds," in Stuart Reifel, ed., *Theory in Context and Out* (Westport, CT: Ablex, 2001), pp. 233–54, and "Just Through the Looking Glass: Children's Understanding of Pretense," in Robert Mitchell, ed., *Pretending and Imagination in Animals and Children* (New York: Cambridge, 2002), pp. 102–14; Aaron Meskin and Jonathan Weinberg, "Imagine That!" in Kieran, ed., *Contemporary Debates in Aesthetics and the Philosophy of Art* (Malden, MA: Blackwell, 2006), pp. 222–35; Shaun Nichols, "Imagination and the Puzzles of Iteration," *Analysis*, LXIII (2003): 182–87, "Imagining and Believing: The Promise of a Single Code," *Journal of Aesthetics and Art Criticism*, LXII (2004): 129–39, "Just the Imagination: Why Imagining Doesn't Behave Like Believing," forthcoming in *Mind and Language*; Josef Perner, *Understanding the Representational Mind* (Cambridge: MIT, 1993); Shaun Nichols and Stephen Stitch, *Mindreading* (New York: Oxford, 2003); and Walton, *Mimesis as Make-Believe*, "Metaphor and Prop Oriented Make-Believe," *The European Journal of Philosophy*, I, 1 (1993): 39–57, "Spelunking, Simulation, and Slime: On Being Moved by Fiction," in Mette and Laver, eds., and "In Other Shoes: The Varieties of Empathy and Their Uses," in his *In Other Shoes* (New York: Oxford, forthcoming).

[22] Our more sophisticated responses to fiction might involve taking the narrator to be unreliable or pretending that the text is something other than a factual narrative. These responses, however, seem parasitic upon a more basic and more naive response to fiction in which we pretend that narrators are describing the real world and where we treat them as infallible authorities concerning what is true in the worlds they describe.

suggests that, in one way or another, a great deal of literary criticism involves this sort of thought or discourse. We enter inside the world of the text being discussed where fictional characters are flesh and blood people while simultaneously adopting a perspective from which we can recognize that the characters we are talking about are fictional objects and are modeled on, or used as the models for, other fictional objects in other texts. Our engagement in literary criticism naturally and seamlessly involves participation in the sorts of pretense or make-believe involved in reading and imaginatively engaging with a fictional text. Criticism is not a practice that is external to and separable from our ordinary imaginative engagement with fiction. Rather it presupposes and incorporates that engagement.

Because of this I think we should regard our talk of fictional characters as simply another sort of pretense which builds upon and extends those pretenses into which we enter in the course of our ordinary imaginative engagement with fictional texts. I will call the sorts of pretenses involved in our ordinary imaginative engagement with fictional texts P_1 *pretenses* and I will call the sorts of pretenses involved in our thought and talk about fictional characters P_2 *pretenses*. On this picture, then, a P_2 pretense involves our engaging in a P_1 pretense in which we pretend that the world is as it is portrayed in the relevant text and, in particular, that it contains various individuals who have the sorts of properties ascribed to them by that text. But, in addition to this, it will also involve us pretending that these individuals have the sorts of properties which fictional realists attribute to fictional characters. We should, in short, pretend that these individuals have a dual nature.[23]

Thus, for example, a P_2 pretense might involve our pretending that Holmes and Watson were human beings born of human parents who lived in Victorian England. But it would also involve our simultaneously pretending that Watson and Holmes are fictional characters which were created by Conan Doyle, are more or less realistic, were modeled in various ways upon various real people and fictional characters, were the inspiration for other fictional characters, and so on. When engaged in this P_2 pretense we would pretend that Holmes and Watson had a dual nature. We would pretend they are as they are described in the Conan Doyle stories. But we would also pretend that they have the sorts of characteristics the fictional realist attributes to fictional objects.[24]

---

[23] This is, of course, something of an oversimplification. First, some fictional texts explicitly acknowledge the fictional status of their characters and so our ordinary engagement with these texts immediately involves us in a P_2 pretense. Second, as a result of our engaging in a P_2 pretense, we may gain new insights into the world of the relevant story and so come to alter the details of the underlying P_1 pretense. For our purposes, however, we may put these complications aside.

We can make this picture more precise as follows. Now exactly what counts true within a given pretense will be determined by a set of basic principles of generation. These will typically include certain foundational claims which are held true within the pretense. But they may also include various specific bridging principles, principles which tell us how what is true within our pretense depends upon what is really true, and there may be specific conditional principles which tell us that if certain things count as being true within our pretense then certain other things will also count as being true.[25] Thus, for example, if we are pretending that we are bakers and that mud pies are cakes then our foundational principles will include the principle that we are to pretend that we are bakers and that mud pies are cakes. Our bridging principles will include the principle that if some participant in our pretense has n mud pies then they are to count as having n cakes. And our conditional principle might include the principle that if anyone counts as having more than fifteen cakes in our pretense then they count as being a millionaire. Finally, let us say that a pretense β *extends* a pretense α just in case all the principles of generation for α are principles of generation for β.

I suggest we may extend any ordinary P_1 pretense to a P_2 pretense by adding some further principles of generation. Now I think that any P_2 pretense must include at least the following conditional principles in order to count as a P_2 pretense at all:

(P1*) If our pretense concerns a creature a, and if a is not a real thing, then a is a fictional character.

(P2*) If our pretense concerns a and b and if a and b are not real things then a and b count as identical within our pretense iff the fictional character of a = the fictional character of b.

---

[24] In pretending this, we would obviously be pretending something metaphysically and perhaps logically impossible but I argued earlier that we are able to pretend impossible things.

[25] See Walton, *Mimesis as Make-Believe*, pp. 35–43 and pp. 138–91, for an extended discussion of principles of generation. In addition to the principles noted above, our pretense will almost certainly also be governed by some restricted version of the Incorporation Principle and the Recursive Principle discussed by Evans, *The Varieties of Reference* (New York: Oxford, 1982), pp. 354–55. Note however that Evans's Incorporation Principle and his Recursive Principle are in general unlikely to hold without restriction for the sorts of reasons discussed by Walton and due to the sorts of cases of *disparity* discussed by Szabo Gendler in her "On the Relationship between Pretense and Belief."

Depending upon our interests and tastes we may then add various further principles to (P1*) and (P2*) governing when a character is to be counted as the creation of a given author, when a character is to count as realistic or unrealistic, when a character is to count as well developed, as well as when a character is to count as the expression of the author's Id, when a character is to count as occupying a given place in a Levi-Straussian structure, when a character is to count as an embodiment of Foucauldian power relations, and so on.

Note that, given that the world of a story contains x just in case we pretend that x exists when we imaginatively engage with that story, then (P1*) and (P2*) are equivalent to (P1) and (P2). Consequently, given that we cannot talk and think about fictional objects without entering into a P_2 pretense, and given that any P_2 pretense is in part constituted by our adopting (P1*) and (P2*), we should expect (P1) and (P2) to have the near platitudinous status they do.

Of course, there is no reason why a P_2 pretense should only extend a single P_1 pretense. On the contrary, a single P_2 pretense might simultaneously extend a number of different P_1 pretenses associated with a number of different stories. In such a case we might imagine a super-world somehow containing all the different worlds of those different stories with the principles of generation for each story restricted to that world. This super-world would contain all the different creatures from those worlds. We would imagine these creatures to have the dual nature of fictional characters. And by engaging in such a pretense we could compare fictional characters from different stories, we could consider which characters occurred in more than one story, we could note which characters served as models for which others, and so on.

There is obviously much more to say here. But for the moment let us contrast the view articulated above with fictional realism. I argued earlier that the fictional realist seems committed to ontic indeterminacy and to fictional objects which flout the law of noncontradiction and the symmetry of identity. The pretense theorist, in contrast, is not subject to these problems. She does not claim that fictional objects really exist, they are not denizens of the real world. Rather, she claims, we simply pretend that there are such things when we engage in a P_2 pretense. And, unlike reality, what we pretend can be indeterminate, inconsistent, and even flout the laws of identity. We may easily engage in a P_2 pretense in which we pretend there are fictional characters while still leaving it open as to whether fictional character *a* exists or whether character *a* is the same as character *b*. The resulting indeterminacies in the world of the pretense will not spill over into the real world. Likewise, we might balk at the suggestion that two objects *a* and *b* are simultaneously both identical and distinct or that *a* is identical to *b* but that *b* is distinct from *a*. But, as I argued earlier, we can pretend these things. So pretense-theoretic accounts of fictional objects can avoid the difficulties with indeterminacy, inconsistency, and asymmetry of identity, which face fictional realism.

IV. THE ARGUMENT FOR FICTIONAL REALISM

At this point, we must turn to the argument for fictional realism noted earlier. Since that argument turned upon the alleged literal truth of utterances of certain *FOS*s, such as (1a)–(1e), we must deny that such utterances are literally true. And we must explain our mistaken intuitions to the contrary. Before doing this, however, I want to pause and raise a further doubt as to whether the fictional realist can really take *FOS*s to be literally true. For consider the following:

(6) (a) Some nineteenth-century fictional characters dote on their mothers more than any eighteenth-century character does.
(b) Conan Doyle's most famous character, Sherlock Holmes, saves the life of Queen Victoria.

Now, *prima facie*, utterances of both (6a) and (6b) seem true. Indeed, they strike me as equally plausible candidates for literal truth as (1a)–(1e). Unfortunately, it is not immediately clear what the fictional realist should say here. For fictional objects do not literally and straightforwardly have mothers in the sense that you and I do. Nor are they the sorts of things that can literally save the Queen. It seems the fictional realist must allow that, despite our intuitions to the contrary, (6a) and (6b) are not literally true. But in this case, it is not clear why we should take (1a)–(1e) to be literally true either.

If the fictional realist is to avoid this conclusion, she must presumably maintain that there is, after all, a genuine sense in which fictional characters can have mothers and save the Queen. But, since a fictional object cannot bear the ordinary property of having-a-mother or saving-the-Queen in the ordinary way, the fictional realist must either distinguish two different but equally genuine senses in which an object may bear properties, or two different but equally genuine sorts of properties an object might bear. On the first option she will distinguish between the ordinary way in which objects may bear properties (call this *exemplification*) and a special way in which fictional objects may bear the properties ascribed to them in stories (call this *encoding*). Sherlock Holmes will not *exemplify* the property of having-saved-the-Queen but he may nevertheless *encode* that property. Likewise, while fictional characters will not *exemplify* the property of having-a-mother, they will nevertheless *encode* that property.[26] On the second option the fictional realist might distinguish between the ordinary *extra-nuclear* properties of having-a-mother and saving-the-Queen and special *nuclear* correlates of those properties which an object may bear in virtue of being ascribed those ordinary properties in a story. Sherlock Holmes will not bear the extra-nuclear property of having-saved-the-Queen but he might nevertheless bear the *nuclear* correlate of that property. And, while fictional characters will not bear the extra-nuclear property of having a mother, they will nevertheless bear its nuclear correlate.[27]

Now I do not want to consider this approach in detail but I want to point out that it faces a number of serious problems. It depends upon our making an obscure and unexplicated distinction between two sorts of predication or two sorts of property where intuitively there seem no such distinctions. It postulates a potential ambiguity in *FOS*s where intuitively there seems to be no such ambiguity. And consider the following piece of discourse:

(7) Both Oedipus and Freud were devoted to their mothers. But while Freud was a real person, Oedipus is only a fictional character.

I think it is natural to regard (7) as true but the first sentence in (7) seems to predicate exactly the same property in exactly the same way of both Oedipus and Freud. This means that in order for (7) to be true, it must be possible for fictional characters and real people to bear the same properties in the same way. In particular, it must be possible for a fictional character to have a mother in exactly the way that a real person has a mother. And this is a problem for the approach we are considering. For, on that approach, Oedipus merely *encodes* or he merely exemplifies the *nuclear correlate* of the properties he is ascribed in the Sophocles's plays. But Freud exemplifies the extra-nuclear property of having a mother. The coherence of (7) suggests that, in the end, the distinction between encoding and exemplification and the distinction between nuclear and extra-nuclear properties are not tenable. We predicate the same properties of fictional characters that we do of real objects and we predicate these properties in the same way. This appears to leave the fictional realist with no alterative but to admit that, despite our intuitions to the contrary, (6a) and (6b) are not literally true. And this, in turn, must call into question our intuitions that (1a)–(1e) are literally true.

---

[26] See Zalta, *Abstract Objects* and *Intensional Logic and the Metaphysics of Intensionality.* Van Inwagen, "Existence and Ontological Commitment," pp. 145–49, distinguishes between the properties a fictional object *has* and those it *holds* while van Inwagen, "Creatures of Fiction," distinguishes between the properties a fictional object *has* and those *ascribed* to it (van Inwagen treats the latter relation as an unexplicated primitive). In both cases he takes the apparatus of predication to ambiguous between the two relations.

[27] For the distinction between nuclear and extra-nuclear properties, see Parsons, *Nonexistent Objects.* Parsons himself takes ordinary properties such as being-green and being-human to be nuclear properties while he takes logical properties such as existence and being-possible to be extra-nuclear properties. For Parsons fictional objects bear the nuclear properties ascribed to them in the texts where they occur and nuclear correlates of the extra-nuclear properties ascribed to them in those texts. Since I do not think we can make sense of a fictional object genuinely having a mother in the sense that real people have mothers, I would argue that Parsons must take the property of having-a-mother to be an extra-nuclear, rather than a nuclear, property and take fictional texts to ascribe its nuclear-correlate rather than the extra-nuclear property itself.

But why do we think that (1a)–(1e) are literally true if they are not? I think the answer has two components. First, although all talk and thought of fictional characters takes place within a P_2 pretense, we nevertheless use utterances of *FOS*s such as (1a)–(1e) to convey information about the real world. More precisely, a P_2 pretense and the P_1 pretenses it extends will be governed by a complex system of principles of generation, some of which will be bridging principles which tell us that certain things count as true within our P_2 pretense just in case certain other things are really true. Now, if a bridging principle tells us that χ is true in our pretense just in case ψ is really true then the assertion that χ, made within the scope of our pretense, may be used to convey the information that ψ is the case. Taking our mud-pie game as an example, I might exclaim "Little Johnny is eating his cake" in order to tell you that Little Johnny is eating a mud pie. I speak within the pretense in order to convey information about the real world.

In the same way, I think, our talk of fictional characters is used to convey information about the real world. I suspect that, at least in very many cases, we intend to convey various sorts of information about how the author of a story, or those reading that story, imagine the world portrayed by that story. For example, to a very rough approximation, if I say that fictional character α in fiction *f* was the model for fictional character β in fiction *g*, I am saying that during the creative process the author of *g* imagined β to be a certain way (and then described β accordingly) as a result of her imagining α to be a certain way when she read *f*. If I say that the character of γ in fiction *h* is an appalling instance of Orientalism, I am saying that during the creative process the author of *h* imagined γ in a way which reflected the worst sorts of Western stereotypes about the East. And so on. We engage in talk of the fictional characters in *f* in order to convey information about how we or others imagine the denizens of *f* to be and why we imagine them to be the way we do. This is real information about the real world. And utterances of a *FOS* made within the scope of a P_2 pretense will count as true within that pretense just in case they convey truths about the real world.

Now, of course, nothing I have said as yet explains why it is that we take utterances of *FOS*s to be *literally* true. After all, we might recognize that my utterance during the mud-pie game of "Little Johnny is eating his cake" is both true in our pretense and conveys a truth about the real world. Still, we feel no temptation to regard my claim as literally true.

I think there are two pertinent differences between our talk of fictional characters and my warnings about Little Johnny. First, and less important, the primary purpose of our talk of fictional characters is the serious business of conveying information about the real world. This is not the primary purpose of our talk within the mud-pie game. Rather such talk functions primarily as part of the mud-pie game itself and conveys information about what is true in that game to those participating in the game. And, although mudpie-game talk can be co-opted for serious purposes, it is hard to ignore the contrast between those infrequent cases in which we use such talk seriously and our normal employment of it. Noticing this contrast brings home to us that what I am literally saying when I claim that Little Johnny is eating his cake is rather different from the information about the real world that I convey. Hence we can recognize that what I say is not literally true.

The second difference, which is much more important and much more interesting, is this. We judge that my utterance of "Little Johnny is eating his cake" is not literally true because we can distinguish between the claim this utterance makes within the scope of our mud-pie pretense and the information I am using it to convey about the real world. Within the scope of our pretense it makes a claim about cake eating. Outside of that pretense it conveys information about eating mud pies.

Crucially, we do not seem to get an analogous phenomenon in the case of P_2 pretenses. For the information we seek to convey when we utter *FOS*s is primarily information about how the author or readers of a text imagine the world portrayed by that text to be and why they imagine it that way. And there seems no way available for me to describe the content of such imaginings except by adopting the perspective of those engaged in them. I must describe the content of these imaginings from the inside, as it were, talking and thinking as if the objects imagined were real things. There might be creatures who were able to communicate the content of their imaginings externally, simply by describing the neural and physical states underpinning their acts of imagination. But we are not them. If I imagine a tree then the only way I can talk about certain aspects of my imagining—the color, size, and shape that I imagine the tree to have, for example—requires me to describe these features from within the perspective of my imaginative act. I must describe the tree I imagine as if it were before me, in the same way I would describe a real tree.

This is equally true when we try to talk about the imaginative acts of those who write fiction. Suppose we want to describe how Virginia Woolf was influenced by Vita Sackville-West when Woolf wrote her novel *Orlando.* This will require our describing how Woolf's imaginative processes were influenced by the way she perceived Sackville-West. More precisely, we will have to describe how Woolf imagined that the world she described in *Orlando* was real, that Orlando was a real person, and that Orlando had certain of the characteristics Woolf perceived in Sackville-West. Now in order to describe all this we must enter into the world of Woolf's story ourselves, as Woolf herself did when writing it, imagining it to be real. We must talk and think as if Orlando exists and shares certain characteristics with Sackville-West. We must, in short, engage in a P_1 pretense. If we do not do this, if we resolutely stand outside the world of Woolf's story and refuse to imagine Orlando, then it seems impossible to talk about the ways in which Orlando resembles Sackville-West. And it will be impossible to isolate the ways in which Sackville-West influenced Woolf's imaginative processes.

Of course, in order to describe Woolf's imaginative processes when writing Orlando, it is not enough merely to engage in a P_1 pretense and imagine that the world of *Orlando* is real. We must also recognize that Woolf determined which features Orlando has, what happens to Orlando, and so on. And we must recognize that Orlando may be viewed in various ways by different readers. We must, in other words, recognize that the objects we are imagining are the creations of authors and the objects of interpretation for readers. And this involves our extending our initial P_1 pretense to a P_2 pretense.

What I am suggesting, then, is that when we try to describe the information about the real world conveyed by utterances of *FOS*s, we find we cannot describe it except in so far as we enter into a P_2 pretense and talk and think as if there really were fictional characters. We cannot step outside of our P_2 pretenses in order to describe this information from an external standpoint. This does not prevent the information we convey from being about the real world. Our acts of imagining are real things with real properties. It is just that the only way to describe many of the most interesting features of these acts requires our participating in the very sorts of imaginative process we seek to describe. Because of this, I think, there is a crucial difference between our talk within the mud-pie game and our talk of fictional characters. In the former case, we are able to clearly distinguish between the claim I made about Little Johnny within the scope of the pretense and the information about the real world conveyed by my utterance for we are able to characterize this information from a perspective external to our pretense. In the case of our talk about fictional characters, this is simply not the case. It is primarily because of this, I claim, that we mistake utterances of *FOSs* such as (1a)–(1e) for literal truths.

I want to conclude by briefly considering how the arguments I have presented against fictional realism are related to the objections leveled by Russell against Meinong's Theory of Objects. Now it is not always easy to follow the course of the Russell-Meinong debate and I do not intend to enter into the niceties of Russell exegesis here. But I think the objections presented above are in the general spirit of Russell's objections though they differ from them in detail and are more general, applying not merely to Meinongian objects but to fictional objects in general.

Russell is concerned to argue that Meinong's objects lead to various violations of the law of noncontradiction and he presents at least two arguments to this effect. The first of these is that, since Meinong is committed to there being a nonexistent object which satisfies the description "the existing King of France," and since presumably the existing King of France exists, Meinong is committed both to the existent king of France existing and to the existent king of France not existing.[28] The second objection is that Meinong is committed to the existence of contradictory objects, such as round squares, for he is committed to there being an object which satisfies the description "the existing round square" and this object must be existent and round and square.[29] It has become customary for Meinongians to try and avoid Russell's objections by distinguishing between two sorts of properties or two forms of predication. Thus, as we have seen, some distinguish between the normal form of predication (exemplification) and a special form of predication (encoding). These theorists hold that while the existent King of France and the existent round square encode the property of existing they do not genuinely exist because they fail to exemplify it.[30] Others, as we have seen, distinguish between so-called nuclear and extra-nuclear properties. And these theorists hold that while the existent King of France and the existent round square have the nuclear property of existing they do not genuinely exist because they lack the extra-nuclear property of existing.[31] Now I do not intend to evaluate these responses to Russell here though I have already expressed some doubts as to the tenability of the exemplification/encoding distinction and the nuclear/extra-nuclear distinction. Rather I simply want to point out that the objections I have raised to fictional realism cannot straightforwardly be blocked by invoking these sorts of distinctions. If my objections are correct, then, the present-day followers of Meinong, as well as other fictional realists, are in worse shape than commonly supposed.

---

[28] See Russell, "On Denoting," *Mind*, XIV, 56 (1905): 479–93, especially p. 438. See also his critical notice of Meinong's *Über die Stellung der Gegenstandstheorie im System der Wissenschaften*, *Mind*, XVI (1907): 436–38.

[29] See Russell, "On Denoting," p. 438, and his critical notice of Meinong's *Gegenstandstheorie und Psychologie*, *Mind*, XIV (1905): 530–38, especially p. 533. See also Nicholas Griffin, "Russell's Critique of Meinong's Theory of Objects," *Grazer Philosophische Studien*, XXV/XXVI (1985/1986): 375–401.

ANTHONY EVERETT

University of Bristol

---

[30] This, basically, is the strategy of Zalta in his *Intensional Logic and the Metaphysics of Intensionality*.

[31] This, basically, is the strategy of Parsons in his *Nonexistent Objects*.

INDEX TO VOLUME CII

INDEX OF SUBJECTS DISCUSSED IN THE JOURNAL

Action, Human

Hieronymi, Pamela: The Wrong Kind of Reason. **102** (9): 437–57 '05

Tomkow, Terrance: The Dif. By Kadri Vilhvelin and Terrance Tomkow. **102** (4): 183–205 '05

Vilhvelin, Kadri: The Dif. By Kadri Vihvelin and Terrance Tomkow. **102** (4): 183–205 '05

Aesthetics

Derksen, Anthony A.: Linear Perspective as a Realist Constraint. **102** (5): 235–58 '05

Everett, Anthony: Against Fictional Realism. **102** (12): 624–48 '05

Westerhoff, Jan: Logical Relations between Pictures. **102** (12): 603–23 '05

Alternative Possibilities, Principle of

Palmer, David: New Distinctions, Same Troubles: A Reply to Haji and McKenna. **102** (9): 474–82 '05

Artificial Intelligence

Churchland, Paul M.: Functionalism at Forty: A Critical Retrospective. **102** (1): 33–50 '05

Belief Reports

Almog, Joseph: Is a Unified Description of Language-and-Thought Possible? **102** (10): 493–531 '05

Ostertag, Gary: A Puzzle about Disbelief. **102** (11): 573–93 '05

Biology, Philosophy of

Andreasen, Robin O.: The Meaning of 'Race': Folk Conceptions and the New Biology of Race. **102** (2): 94–106 '05

LaPorte, Joseph: Is There a Single Objective, Evolutionary Tree of Life? **102** (7): 357–74 '05

Causation

Tomkow, Terrance: The Dif. By Kadri Vilhvelin and Terrance Tomkow. **102** (4): 183–205 '05

Vilhvelin, Kadri: The Dif. By Kadri Vihvelin and Terrance Tomkow. **102** (4): 183–205 '05

Choice Theory, Rational

Laden, Anthony Simon: Evaluating Social Reasons: Hobbes versus Hegel. **102** (7): 327–56 '05

Tungodden, Bertil: On the Possibility of Paretian Egalitarianism. By Bertil Tungodden and Peter Vallentyne. **102** (3): 126–54 '05

Vallentyne, Peter: On the Possibility of Paretian Egalitarianism. By Bertil Tungodden and Peter Vallentyne. **102** (3): 126–54 '05

Class Membership

Fine, Kit: Class and Membership. **102** (11): 547–72 '05

Color

Averill, Edward Wilson: Toward a Projectivist Account of Color. **102** (5): 217–34 '05

Consequentialism

Tiberius, Valerie: Wisdom and Perspective. **102** (4): 163–82 '05

Tomkow, Terrance: The Dif. By Kadri Vilhvelin and Terrance Tomkow. **102** (4): 183–205 '05

Vilhvelin, Kadri: The Dif. By Kadri Vihvelin and Terrance Tomkow. **102** (4): 183–205 '05

Counterfactuals

Tomkow, Terrance: The Dif. By Kadri Vilhvelin and Terrance Tomkow. **102** (4): 183–205 '05

Vilhvelin, Kadri: The Dif. By Kadri Vihvelin and Terrance Tomkow. **102** (4): 183–205 '05

Deception

Strudler, Alan: Deception Unraveled. **102** (9): 458–82 '05

Depiction

Derksen, Anthony A.: Linear Perspective as a Realist Constraint. **102** (5): 235–58 '05

Westerhoff, Jan: Logical Relations between Pictures. **102** (12): 603–23 '05

Descriptions, Theory of

Almog, Joseph: Is a Unified Description of Language-and-Thought Possible? **102** (10): 493–531 '05

Determinism

Palmer, David: New Distinctions, Same Troubles: A Reply to Haji and McKenna. **102** (9): 474–82 '05

Domsky, Darren

Statman, Daniel: Doors, Keys, and Moral Luck: A Reply to Domsky. **102** (8): 422–36 '05

Economics

Tungodden, Bertil: On the Possibility of Paretian Egalitarianism. By Bertil Tungodden and Peter Vallentyne. **102** (3): 126–54 '05

Vallentyne, Peter: On the Possibility of Paretian Egalitarianism. By Bertil Tungodden and Peter Vallentyne. **102** (3): 126–54 '05

Egalitarianism

Tungodden, Bertil: On the Possibility of Paretian Egalitarianism. By Bertil Tungodden and Peter Vallentyne. **102** (3): 126–54 '05

Vallentyne, Peter: On the Possibility of Paretian Egalitarianism. By Bertil Tungodden and Peter Vallentyne. **102** (3): 126–54 '05

Epistemology

See Knowledge, Theory of

Ethics

Ernst, Zachary: A Plea for Asymmetric Games. **102** (3): 109–25 '05

Hieronymi, Pamela: The Wrong Kind of Reason. **102** (9): 437–57 '05

Laden, Anthony Simon: Evaluating Social Reasons: Hobbes versus Hegel. **102** (7): 327–56 '05

Palmer, David: New Distinctions, Same Troubles: A Reply to Haji and McKenna. **102** (9): 474–82 '05

Statman, Daniel: Doors, Keys, and Moral Luck: A Reply to Domsky. **102** (8): 422–36 '05

Strudler, Alan: Deception Unraveled. **102** (9): 458–82 '05

Tiberius, Valerie: Wisdom and Perspective. **102** (4): 163–82 '05

Tomkow, Terrance: The Dif. By Kadri Vilhvelin and Terrance Tomkow. **102** (4): 183–205 '05

Tungodden, Bertil: On the Possibility of Paretian Egalitarianism. By Bertil Tungodden and Peter Vallentyne. **102** (3): 126–54 '05

Vallentyne, Peter: On the Possibility of Paretian Egalitarianism. By Bertil Tungodden and Peter Vallentyne. **102** (3): 126–54 '05

Vilhvelin, Kadri: The Dif. By Kadri Vihvelin and Terrance Tomkow. **102** (4): 183–205 '05

Ethics, Virtue

Tiberius, Valerie: Wisdom and Perspective. **102** (4): 163–82 '05

Evolution

Andreasen, Robin O.: The Meaning of 'Race': Folk Conceptions and the New Biology of Race. **102** (2): 94–106 '05

Ernst, Zachary: A Plea for Asymmetric Games. **102** (3): 109–25 '05

LaPorte, Joseph: Is There a Single Objective, Evolutionary Tree of Life? **102** (7): 357–74 '05

Existence

Efird, David: The Subtraction Argument for Metaphysical Nihilism. By

David Efird and Tom Stoneham. **102** (6): 303–25 '05

Everett, Anthony: Against Fictional Realism. **102** (12): 624–48 '05

Koslicki, Kathrin: Almost Indiscernible Objects and the Suspect Strategy. **102** (2): 55–77 '05

Smith, Nicholas J.J.: A Plea for Things That Are Not Quite All There: Or, Is There a Problem about Vague Composition and Vague Existence? **102** (8): 381–421 '05

Stoneham, Tom: The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 '05

Explanation

Churchland, Paul M.: Functionalism at Forty: A Critical Retrospective. **102** (1): 33–50 '05

Ernst, Zachary: A Plea for Asymmetric Games. **102** (3): 109–25 '05

Johnsen, Bredo C.: How to Read "Epistemology Naturalized." **102** (2): 78–93 '05

Schwartz, Robert: A Note on Goodman's Problem. **102** (7): 375–79 '05

Frankfurt, Harry

Palmer, David: New Distinctions, Same Troubles: A Reply to Haji and McKenna. **102** (9): 474–82 '05

Free Will

Palmer, David: New Distinctions, Same Troubles: A Reply to Haji and McKenna. **102** (9): 474–82 '05

Frege, Gottlob

Almog, Joseph: Is a Unified Description of Language-and-Thought Possible? **102** (10): 493–531 '05

Gallois, André: The Simplicity of Identity. **102** (6): 273–302 '05

Functionalism

Churchland, Paul M.: Functionalism at Forty: A Critical Retrospective. **102** (1): 33–50 '05

Game Theory

Ernst, Zachary: A Plea for Asymmetric Games. **102** (3): 109–25 '05

Genetics

Andreasen, Robin O.: The Meaning of 'Race': Folk Conceptions and the New Biology of Race. **102** (2): 94–106 '05

LaPorte, Joseph: Is There a Single Objective, Evolutionary Tree of Life? **102** (7): 357–74 '05

Good Life

Tiberius, Valerie: Wisdom and Perspective. **102** (4): 163–82 '05

Goodman, Nelson

Schwartz, Robert: A Note on Goodman's Problem. **102** (7): 375–79 '05

Glasgow, Joshua

Andreasen, Robin O.: The Meaning of 'Race': Folk Conceptions and the New Biology of Race. **102** (2): 94–106 '05

Grue

Schwartz, Robert: A Note on Goodman's Problem. **102** (7): 375–79 '05

Haji, Ishtiyaque

Palmer, David: New Distinctions, Same Troubles: A Reply to Haji and McKenna. **102** (9): 474–82 '05

Hegel, G.W.F.

Laden, Anthony Simon: Evaluating Social Reasons: Hobbes versus Hegel. **102** (7): 327–56 '05

Hobbes, Thomas

Laden, Anthony Simon: Evaluating Social Reasons: Hobbes versus Hegel. **102** (7): 327–56 '05

Identity Theory

Efird, David: The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 '05

Gallois, André: The Simplicity of Identity. **102** (6): 273–302 '05

Koslicki, Kathrin: Almost Indiscernible Objects and the Suspect Strategy. **102** (2): 55–77 '05

Smith, Nicholas J.J.: A Plea for Things That Are Not Quite All There: Or, Is There a Problem about Vague Composition and Vague Existence? **102** (8): 381–421 '05

Stoneham, Tom: The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 '05

Indiscernibles, Identity of

Efird, David: The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 '05

Hofweber, Thomas: Supervenience and Object-dependent Properties. **102** (1): 5–32 '05

Koslicki, Kathrin: Almost Indiscernible Objects and the Suspect Strategy. **102** (2): 55–77 '05

Stoneham, Tom: The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 '05

Justice

Laden, Anthony Simon: Evaluating Social Reasons: Hobbes versus Hegel. **102** (7): 327–56 '05

Tungodden, Bertil: On the Possibility of Paretian Egalitarianism. By Bertil Tungodden and Peter Vallentyne. **102** (3): 126–54 '05

Vallentyne, Peter: On the Possibility of Paretian Egalitarianism. By Bertil Tungodden and Peter Vallentyne. **102** (3): 126–54 '05

Knowledge, Theory of

Armour-Garb, Bradley: Standing on Common Ground. By Jody Azzouni and Bradley Armour-Garb. **102** (10): 532–44 '05

Azzouni, Jody: Standing on Common Ground. By Jody Azzouni and Bradley Armour-Garb. **102** (10): 532–44 '05

Gallois, André: The Simplicity of Identity. **102** (6): 273–302 '05

Johnsen, Bredo C.: How to Read "Epistemology Naturalized." **102** (2): 78–93 '05

Ostertag, Gary: A Puzzle about Disbelief. **102** (11): 573–93 '05

Schwartz, Robert: A Note on Goodman's Problem. **102** (7): 375–79 '05

Koslow, Arnold

Westerhoff, Jan: Logical Relations between Pictures. **102** (12): 603–23 '05

Language, Philosophy of

Almog, Joseph: Is a Unified Description of Language-and-Thought Possible? **102** (10): 493–531 '05

Gallois, André: The Simplicity of Identity. **102** (6): 273–302 '05

Hofweber, Thomas: Supervenience and Object-dependent Properties. **102** (1): 5–32 '05

Ostertag, Gary: A Puzzle about Disbelief. **102** (11): 573–93 '05

Leibniz's Law

Gallois, André: The Simplicity of Identity. **102** (6): 273–302 '05

Koslicki, Kathrin: Almost Indiscernible Objects and the Suspect Strategy. **102** (2): 55–77 '05

Lewis, David

Armour-Garb, Bradley: Standing on Common Ground. By Jody Azzouni and Bradley Armour-Garb. **102** (10): 532–44 '05

Azzouni, Jody: Standing on Common Ground. By Jody Azzouni and Bradley Armour-Garb. **102** (10): 532–44 '05

Literature

Everett, Anthony: Against Fictional Realism. **102** (12): 624–48 '05

Logic, Philosophy of

Almog, Joseph: Is a Unified Description of Language-and-Thought Possible? **102** (10): 493–531 '05

Armour-Garb, Bradley: Standing on Common Ground. By Jody Azzouni and Bradley Armour-Garb. **102** (10): 532–44 '05

Azzouni, Jody: Standing on Common Ground. By Jody Azzouni and Bradley Armour-Garb. **102** (10): 532–44 '05

Fine, Kit: Class and Membership. **102** (11): 547–72 '05

Smith, Nicholas J.J.: A Plea for Things That Are Not Quite All There: Or, Is There a Problem about Vague Composition and Vague Existence? **102** (8): 381–421 '05

Westerhoff, Jan: Logical Relations between Pictures. **102** (12): 603–23 '05

Luck, Moral

Statman, Daniel: Doors, Keys, and Moral Luck: A Reply to Domsky. **102** (8): 422–36 '05

Mathematics, Philosophy of

Fine, Kit: Class and Membership. **102** (11): 547–72 '05

McKenna, Michael

Palmer, David: New Distinctions, Same Troubles: A Reply to Haji and McKenna. **102** (9): 474–82 '05

Memo to Authors

102 (1) 51 '05

Metaphysics

Efird, David: The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 '05

Everett, Anthony: Against Fictional Realism. **102** (12): 624–48 '05

Gallois, André: The Simplicity of Identity. **102** (6): 273–302 '05

Hofweber, Thomas: Supervenience and Object-dependent Properties. **102** (1): 5–32 '05

Koslicki, Kathrin: Almost Indiscernible Objects and the Suspect Strategy. **102** (2): 55–77 '05

Smith, Nicholas J.J.: A Plea for Things That Are Not Quite All There: Or, Is There a Problem about Vague Composition and Vague Existence? **102** (8): 381–421 '05

Stoneham, Tom: The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 '05

Method, Scientific

Johnsen, Bredo C.: How to Read "Epistemology Naturalized." **102** (2): 78–93 '05

Mind, Philosophy of

Averill, Edward Wilson: Toward a Projectivist Account of Color. **102** (5): 217–34 '05

Churchland, Paul M.: Functionalism at Forty: A Critical Retrospective. **102** (1): 33–50 '05

Westerhoff, Jan: Logical Relations between Pictures. **102** (12): 603–23 '05

Multiple Realizability Thesis

Churchland, Paul M.: Functionalism at Forty: A Critical Retrospective. **102** (1): 33–50 '05

Neuroscience

Averill, Edward Wilson: Toward a Projectivist Account of Color. **102** (5): 217–34 '05

Churchland, Paul M.: Functionalism at Forty: A Critical Retrospective. **102** (1): 33–50 '05

New Books

102 (1): 52–54 '05
102 (2): 108 '05
102 (9): 487–92 '05
102 (10): 545–46 '05
102 (11): 602 '05

New Books: Translations

102 (6): 326 '05

Nihilism

Efird, David: The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 '05

Stoneham, Tom: The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 '05

Noncontradiction, Law of

Armour-Garb, Bradley: Standing on Common Ground. By Jody Azzouni and Bradley Armour-Garb. **102** (10): 532–44 '05

Azzouni, Jody: Standing on Common Ground. By Jody Azzouni and Bradley Armour-Garb. **102** (10): 532–44 '05

Notes and News

102 (2): 107 '05

Perception

Averill, Edward Wilson: Toward a Projectivist Account of Color. **102** (5): 217–34 '05

Derksen, Anthony A.: Linear Perspective as a Realist Constraint. **102** (5): 235–58 '05

Perspective, Linear

Derksen, Anthony A.: Linear Perspective as a Realist Constraint. **102** (5): 235–58 '05

Physicalism

Hofweber, Thomas: Supervenience and Object-dependent Properties. **102** (1): 5–32 '05

Political Theory

Laden, Anthony Simon: Evaluating Social Reasons: Hobbes versus Hegel. **102** (7): 327–56 '05

Psychology

Churchland, Paul M.: Functionalism at Forty: A Critical Retrospective. **102** (1): 33–50 '05

Johnsen, Bredo C.: How to Read "Epistemology Naturalized." **102** (2): 78–93 '05

Statman, Daniel: Doors, Keys, and Moral Luck: A Reply to Domsky. **102** (8): 422–36 '05

Quine, W.V.

Johnsen, Bredo C.: How to Read "Epistemology Naturalized." **102** (2): 78–93 '05

Race

Andreasen, Robin O.: The Meaning of 'Race': Folk Conceptions and the New Biology of Race. **102** (2): 94–106 '05

Rationality

Hieronymi, Pamela: The Wrong Kind of Reason. **102** (9): 437–57 '05

Johnsen, Bredo C.: How to Read "Epistemology Naturalized." **102** (2): 78–93 '05

Laden, Anthony Simon: Evaluating Social Reasons: Hobbes versus Hegel. **102** (7): 327–56 '05

Tiberius, Valerie: Wisdom and Perspective. **102** (4): 163–82 '05

Realism

Derksen, Anthony A.: Linear Perspective as a Realist Constraint. **102** (5): 235–58 '05

Everett, Anthony: Against Fictional Realism. **102** (12): 624–48 '05

Reasons

Hieronymi, Pamela: The Wrong Kind of Reason. **102** (9): 437–57 '05

Laden, Anthony Simon: Evaluating Social Reasons: Hobbes versus Hegel. **102** (7): 327–56 '05

Reference

Almog, Joseph: Is a Unified Description of Language-and-Thought Possible? **102** (10): 493–531 '05

Everett, Anthony: Against Fictional Realism. **102** (12): 624–48 '05

Gallois, André: The Simplicity of Identity. **102** (6): 273–302 '05

Hofweber, Thomas: Supervenience and Object-dependent Properties. **102** (1): 5–32 '05

Ostertag, Gary: A Puzzle about Disbelief. **102** (11): 573–93 '05

Representation, Visual

Averill, Edward Wilson: Toward a Projectivist Account of Color. **102** (5): 217–34 '05

Westerhoff, Jan: Logical Relations between Pictures. **102** (12): 603–23 '05

Responsibility

Statman, Daniel: Doors, Keys, and Moral Luck: A Reply to Domsky. **102** (8): 422–36 '05

Strudler, Alan: Deception Unraveled. **102** (9): 458–82 '05

Tomkow, Terrance: The Dif. By Kadri Vilhvelin and Terrance Tomkow. **102** (4): 183–205 '05

Vilhvelin, Kadri: The Dif. By Kadri Vihvelin and Terrance Tomkow. **102** (4): 183–205 '05

Science, Philosophy of

Andreasen, Robin O.: The Meaning of 'Race': Folk Conceptions and the New Biology of Race. **102** (2): 94–106 '05

Averill, Edward Wilson: Toward a Projectivist Account of Color. **102** (5): 217–34 '05

Churchland, Paul M.: Functionalism at Forty: A Critical Retrospective. **102** (1): 33–50 '05

LaPorte, Joseph: Is There a Single Objective, Evolutionary Tree of Life? **102** (7): 357–74 '05

Schwartz, Robert: A Note on Goodman's Problem. **102** (7): 375–79 '05

Self-knowledge

Hieronymi, Pamela: The Wrong Kind of Reason. **102** (9): 437–57 '05

Semantics

Almog, Joseph: Is a Unified Description of Language-and-Thought Possible? **102** (10): 493–531 '05

Gallois, André: The Simplicity of Identity. **102** (6): 273–302 '05

Hofweber, Thomas: Supervenience and Object-dependent Properties. **102** (1): 5–32 '05

Koslicki, Kathrin: Almost Indiscernible Objects and the Suspect Strategy. **102** (2): 55–77 '05

Ostertag, Gary: A Puzzle about Disbelief. **102** (11): 573–93 '05

Schwartz, Robert: A Note on Goodman's Problem. **102** (7): 375–79 '05

Set Theory

Fine, Kit: Class and Membership. **102** (11): 547–72 '05

Smith, Nicholas J.J.: A Plea for Things That Are Not Quite All There: Or, Is There a Problem about Vague Composition and Vague Existence? **102** (8): 381–421 '05

Social Theory

Andreasen, Robin O.: The Meaning of 'Race': Folk Conceptions and the New Biology of Race. **102** (2): 94–106 '05

Ernst, Zachary: A Plea for Asymmetric Games. **102** (3): 109–25 '05

Laden, Anthony Simon: Evaluating Social Reasons: Hobbes versus Hegel. **102** (7): 327–56 '05

Tungodden, Bertil: On the Possibility of Paretian Egalitarianism. By Bertil Tungodden and Peter Vallentyne. **102** (3): 126–54 '05

Vallentyne, Peter: On the Possibility of Paretian Egalitarianism. By Bertil Tungodden and Peter Vallentyne. **102** (3): 126–54 '05

Supervenience

Hofweber, Thomas: Supervenience and Object-dependent Properties. **102** (1): 5–32 '05

Trust

Strudler, Alan: Deception Unraveled. **102** (9): 458–82 '05

Utilitarianism

Tiberius, Valerie: Wisdom and Perspective. **102** (4): 163–82 '05

Vagueness

Efird, David: The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 '05

Gallois, André: The Simplicity of Identity. **102** (6): 273–302 '05

Smith, Nicholas J.J.: A Plea for Things That Are Not Quite All There: Or, Is There a Problem about Vague Composition and Vague Existence? **102** (8): 381–421 '05

Stoneham, Tom: The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 '05

Values, Theory of

Hieronymi, Pamela: The Wrong Kind of Reason. **102** (9): 437–57 '05

Tiberius, Valerie: Wisdom and Perspective. **102** (4): 163–82 '05

Vision

Averill, Edward Wilson: Toward a Projectivist Account of Color. **102** (5): 217–34 '05

Derksen, Anthony A.: Linear Perspective as a Realist Constraint. **102** (5): 235–58 '05

Wisdom

Tiberius, Valerie: Wisdom and Perspective. **102** (4): 163–82 '05

Worlds, Possible

Efird, David: The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 '05

Stoneham, Tom: The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 '05

INDEX OF AUTHORS
Including Contributing Authors and Reviewers

Almog, Joseph, Is a Unified Description of Language-and-Thought Possible? **102** (10): 493–531 '05

Andreasen, Robin O., The Meaning of 'Race': Folk Conceptions and the New Biology of Race. **102** (2): 94–106 '05

Armour-Garb, Bradley, Standing on Common Ground. By Jody Azzouni and Bradley Armour-Garb. **102** (10): 532–44 '05

Averill, Edward Wilson, Toward a Projectivist Account of Color. **102** (5): 217–34 '05

Azzouni, Jody, Standing on Common Ground. By Jody Azzouni and Bradley Armour-Garb. **102** (10): 532–44 '05

Block, Ned, Review of Alva Noe: *Action in Perception.* **102** (5): 259–72 '05

Churchland, Paul M., Functionalism at Forty: A Critical Retrospective. **102** (1): 33–50 '05

Derksen, Anthony A., Linear Perspective as a Realist Constraint. **102** (5): 235–58 '05

Efird, David, The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 '05

Ernst, Zachary, A Plea for Asymmetric Games. **102** (3): 109–25 '05

Everett, Anthony, Against Fictional Realism. **102** (12): 624–48 '05

Fine, Kit, Class and Membership. **102** (11): 547–72 '05

Gallois, André, The Simplicity of Identity. **102** (6): 273–302 '05

Hieronymi, Pamela, The Wrong Kind of Reason. **102** (9): 437–57 ’05

Hitz, Zena, Review of Kurt Raaflaub: *The Discovery of Freedom in Ancient Greece.* **102** (11): 594–601 ’05

Hofweber, Thomas, Supervenience and Object-dependent Properties. **102** (1): 5–32 ’05

Johnsen, Bredo C., How to Read “Epistemology Naturalized.” **102** (2): 78–93 ’05

Koslicki, Kathrin, Almost Indiscernible Objects and the Suspect Strategy. **102** (2): 55–77 ’05

Laden, Anthony Simon, Evaluating Social Reasons: Hobbes versus Hegel. **102** (7): 327–56 ’05

LaPorte, Joseph, Is There a Single Objective, Evolutionary Tree of Life? **102** (7): 357–74 ’05

Matthen, Mohan, Review of Philip Kitcher: *In Mendel’s Mirror: Philosophical Reflections on Biology.* **102** (4): 206–16 ’05

Ostertag, Gary, A Puzzle about Disbelief. **102** (11): 573–93 ’05

Palmer, David, New Distinctions, Same Troubles: A Reply to Haji and McKenna. **102** (9): 474–82 ’05

Priest, Graham, Review of Tim Maudlin: *Truth and Paradox: Solving Riddles.* **102** (9): 483–86 ’05

Schwartz, Robert, A Note on Goodman’s Problem. **102** (7): 375–79 ’ 05

Smith, Nicholas J.J., A Plea for Things That Are Not Quite All There: Or, Is There a Problem about Vague Composition and Vague Existence? **102** (8): 381–421 ’05

Statman, Daniel, Doors, Keys, and Moral Luck: A Reply to Domsky. **102** (8): 422–36 ’05

Stoneham, Tom, The Subtraction Argument for Metaphysical Nihilism. By David Efird and Tom Stoneham. **102** (6): 303–25 ’05

Strudler, Alan, Deception Unraveled. **102** (9): 458–82 ’05

Tennant, Neil, Review of Christopher Peacocke: *The Realm of Reason.* **102** (3): 155–62 ’05

Tiberius, Valerie, Wisdom and Perspective. **102** (4): 163–82 ’05

Tomkow, Terrance, The Dif. By Kadri Vilhvelin and Terrance Tomkow. **102** (4): 183–205 ’05

Tungodden, Bertil, On the Possibility of Paretian Egalitarianism. By Bertil Tungodden and Peter Vallentyne. **102** (3): 126–54 ’05

Vallentyne, Peter, On the Possibility of Paretian Egalitarianism. By Bertil Tungodden and Peter Vallentyne. **102** (3): 126–54 ’05

Vilhvelin, Kadri, The Dif. By Kadri Vihvelin and Terrance Tomkow. **102** (4): 183–205 ’05

Westerhoff, Jan, Logical Relations between Pictures. **102** (12): 603–23 ’05

INDEX OF BOOKS REVIEWED
Listed by Author of Item Reviewed

Kitcher, Philip: *In Mendel’s Mirror: Philosophical Reflections on Biology*—Mohan Matthen. **102** (4): 205–16 ’05

Maudlin, Tim: *Truth and Paradox*—Graham Priest. **102** (9): 483–86 ’05

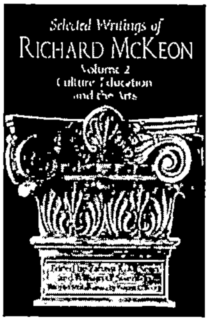
Noe, Alva: *Action in Perception*—Ned Block. **102** (5) 259–77 ’05

Peacocke, Christopher: *The Realm of Reason*—Neil Tennant. **102** (3): 155–62 ’05

Raaflaub, Kurt: *The Discovery of Freedom in Ancient Greece*—Zena Hitz. **102** (11): 594–601 ’05

PHILOSOPHY

NEW FROM CHICAGO



SELECTED WRITINGS OF RICHARD MCKEON, VOLUME TWO

Culture, Education, and the Arts

Richard P. McKeon

Edited by Zahava K. McKeon and William G. Swenson

Introduction by Wayne C. Booth.

Richard McKeon was a philosopher of extraordinary creativity who brought profoundly original ideas to bear on more standard ways of thinking and learning. A classicist, medievalist, and revolutionary intellectual, he fashioned an approach to philosophy as a plural conversation among varied traditions of thought, epochs, and civilizations. This second volume of McKeon's selected works demonstrates his approach to inquiry and practice in culture, education, and the arts.

CLOTH $50.00

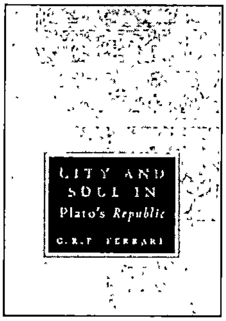
SUSTAINABILITY

A Philosophy of Adaptive Ecosystem Management

Bryan G. Norton

"*Sustainability* is a stunning achievement. It gives direction and purpose to adaptive management, as well as to the environmental sciences and professions that support community efforts to negotiate and achieve environmental sustainability. I suspect it will become a classic."—Bruce Hull, Virginia Tech

PAPER $37.50



Now in paperback

CITY AND SOUL IN PLATO'S *Republic*

G. R. F. Ferrari

"G. R. F. Ferrari's book reopens, from a fresh and original point of view, the long-lasting discussion on the relation between soul and city in the eighth book of Plato's *Republic.* Pointing out its metaphorical rather than interactive character, the book thus confronts the crucial problem of the link between ethics and politics in Plato (which is not denied but rather loosened by Ferrari). It represents, in my opinion, one of the most brilliant and thought-provoking (although debatable) contributions to the renewal of our understanding of Plato's 'masterpiece' and of the overall meaning of his philosophy about man and the city."—Mario Vegetti, University of Pavia

PAPER $17.00

Available this winter from the Center for the Study of Language and Information

UNIVERSAL LOGIC

Ross Brady

LECTURE NOTES

PAPER $35.00

AFTER EUCLID

Jesse Norman

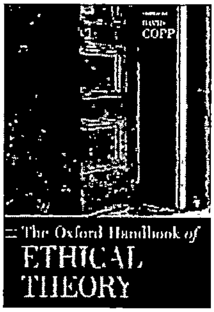
LECTURE NOTES

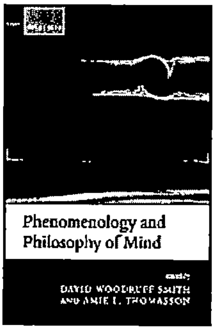
PAPER $25.00

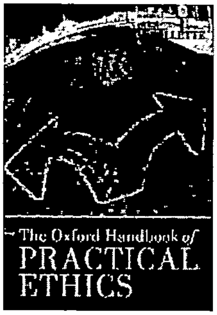
THE UNIVERSITY OF CHICAGO PRESS

1427 East 60th Street, Chicago, IL 60637 • www.press.uchicago.edu

New from
Oxford







Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.us.oup.com/us

OXFORD
UNIVERSITY PRESS

The Oxford Handbook of Ethical Theory

EDITED BY DAVID COPP

The Oxford Handbook of Ethical Theory is a major new reference work in ethical theory consisting of commissioned essays by leading moral philosophers.
(Oxford Handbooks)

November 2005 $74 00

Artha

Meaning

JONARDON GANERI

This second volume in the Foundations of Philosophy in India series is an important contribution to the philosophy of language. Here Jonardon Ganeri highlights the significant relationship between semantic power and epistemic power to understand the important philosophical category of meaning.

November 2005 $29.95

Reflections on Meaning

PAUL HORWICH

Paul Horwich's main aim in *Reflections on Meaning* is to explain how mere noises, marks, gestures, and mental symbols are able to capture the world—that is, how words and sentences (in whatever medium) come to *mean* what they do, to *stand* for certain things, to be true or false of reality. His answer is a groundbreaking development of Wittgenstein's idea that the meaning of a term is nothing more than its *use*.

January 2006 paper $26 95 cloth $74.00

New in Paperback!

The Oxford Handbook of Practical Ethics

Edited by HUGH LAFOLLETTE

The Oxford Handbook of Practical Ethics is a lively and authoritative guide to current thought about ethical issues in all areas of human activity—personal, medical, sexual, political, legal, and beyond.
(Oxford Handbooks)

2003 (paper November 2005) paper $45 00 cloth $195.00

Phenomenology and Philosophy of Mind

EDITED BY DAVID WOODRUFF SMITH AND AMIE L. THOMASSON

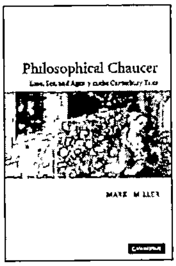
Philosophical work on the mind flowed in two streams through the 20th century: phenomenology and analytic philosophy. This volume aims to bring them together again, by demonstrating how work in phenomenology may lead to significant progress on problems central to current analytic research, and how analytical philosophy of mind may shed light on phenomenological concerns.

November 2005 paper $35 00 cloth $95.00

CAMBRIDGE

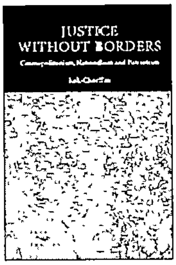
OUTSTANDING SCHOLARSHIP

Philosophical Chaucer
Love, Sex, and Agency in the Canterbury Tales
Mark Miller
Cambridge Studies in Medieval Literature 55
$75.00: Hardback: 0-521-84236-0: 302pp



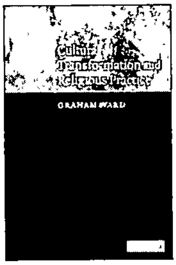
Freedom of Speech
Volume 21: Part 2
Ellen Frankel Paul, Fred D. Miller, Jr., and Jeffrey Paul, Editors
Social Philosophy and Policy
$27.99: Paperback: 0-521-60375-7: 460pp

Justice without Borders
Cosmopolitanism, Nationalism and Patriotism
Kok-Chor Tan
Contemporary Political Theory
$70.00: Hardback: 0-521-83454-6: 234pp
$25.99: Paperback: 0-521-54232-4



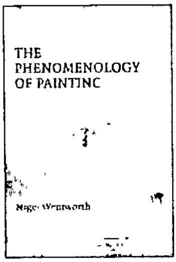
Hardwired Behavior
What Neuroscience Reveals about Morality
Laurence Tancredi
$28.99: Hardback: 0-521-86001-6: 240pp

The Judicial Process
Realism, Pragmatism, Practical Reasoning and Principles
E.W. Thomas
$95 00: Hardback: 0-521-85566-7: 442pp



Phenomenology, Logic, and the Philosophy of Mathematics
Richard Tieszen
$70.00: Hardback: 0-521-83782-0: 368pp

Cultural Transformation and Religious Practice
Graham Ward
$75.00· Hardback: 0-521-83326-4: 194pp
$29.99: Paperback: 0-521-54074-7



Gandhi as Disciple and Mentor
Thomas Weber
$75.00: Hardback: 0-521-84230-1: 294pp

The Phenomenology of Painting
Nigel Wentworth
$80.00: Hardback: 0-521-81999-7: 296pp

Prices subject to change

www.cambridge.org/us/philosophy/

CAMBRIDGE
UNIVERSITY PRESS

new from The MIT Press

Teleological Realism

Mind, Agency, and Explanation
Scott Sehon

"This is one of the strongest recent defenses of the non-causal view, and it deserves to be taken seriously by all philosophers of mind and action."
— Tim Crane, University College London

A Bradford Book • 264 pp. $36

Patient Autonomy and the Ethics of Responsibility

Alfred I. Tauber

"With insight and wide-ranging scholarship, Alfred Tauber analyzes how our reliance on patient autonomy must be tempered by a moral intimacy between the patient and the healer." — Paul Root University of Pennsylvania

Basic Bioethics series • 344 pp. $25 paper

Neither Brain nor Ghost

A Nondualist Alternative to the Mind-Brain Identity Theory
W. Teed Rockwell

"Well researched and well written, this is an excellent introduction to the nascent field of nonlinear neurodynamics."
— Walter J. Freeman, University of California, Berkeley, author of *How Brains Make Up Their Minds*

A Bradford Book • 256 pp., 6 illus. $36

The Harmonic Mind

From Neural Computation to Optimality-Theoretic Grammar
Volume I: Cognitive Architecture
Volume II: Linguistic and Philosophical Implications
Paul Smolensky and Géraldine Legendre

An integrated connectionist/symbolic architecture of the mind/brain, applied to neural/genomic realization of grammar; acquisition, processing, and typology in phonology and syntax; and foundations of cognitive explanation.

A Bradford Book
Volume I: 904 pp. $55
Volume II: 896 pp. $55
2-volume set: 1800 pp. $100

To order call **800-405-1619.**
http://mitpress.mit.edu

Conditionals in Context

Christopher Gauker

"The most sophisticated and comprehensive theory of the semantics and logic of conditionals yet available."
— E. J. Lowe, University of Durham

Contemporary Philosophical Monographs series
A Bradford Book • 344 pp., 2 illus. $35 paper

Law and Social Justice

edited by Joseph Keim Campbell, Michael O'Rourke, and David Shier

Leading scholars consider a variety of philosophical issues in law and social justice, from foundational concepts to specific legal problems.

Topics in Contemporary Philosophy, Volume 3
A Bradford Book • 384 pp., 1 illus. $35 paper

The Primacy of the Subjective

Foundations for a Unified Theory of Mind and Language
Nicholas Georgalis

"Nicholas Georgalis's book is one of the most important contributions to the subject in recent years." — John Searle, University of California, Berkeley

A Bradford Book • 336 pp. $36

Pain

New Essays on Its Nature and the Methodology of Its Study
edited by Murat Aydede

Leading philosophers and scientists examine the puzzles of pain and consider how the study of pain might help us to have a better understanding of the larger issues of consciousness and psychological research.

A Bradford Book • 420 pp., 10 illus. $35 paper

Zone Books

The Order of Evils

Toward an Ontology of Morals
Adi Ophir, translated by Rela Mazali and Havi Carel

"This is the most sustained and brilliant philosophical discussion of evil I have ever read." — Moshe Halbertal, Hebrew University

Distributed for Zone Books • 688 pp. $38.50

now in paperback

Truth and Justification

Jürgen Habermas
translated by Barbara Fultner

"Habermas confronts the writings and insights of his contemporaries and uses them to deepen his own understanding of what it is that philosophy must be about in our own time." — *Bridges*

Studies in Contemporary German Social Thought series • 352 pp. $20 paper

Adorno's Negative Dialectic

Philosophy and the Possibility of Critical Rationality

Brian O'Connor

"Provides a much needed and exceedingly lucid treatment of Adorno's central concerns with the nature of the object of experience and the shape of subjectivity." — Tom Huhn, School of Visual Arts, New York

Studies in Contemporary German Social Thought series • 224 pp. $18 paper

Critical Resistance

From Poststructuralism to Post-Critique

David Couzens Hoy

"There is no better guide through the thickets of poststructualism and its aftermath." — Thomas McCarthy, Northwestern University

288 pp. $18 paper

Real Natures and Familiar Objects

Crawford L. Elder

"Elder defends the much maligned 'ordinary objects' of common sense with rigor and detail against such opponents as reductionists and soriteans."
— Richard Grandy, Rice University

A Bradford Book • 224 pp. $18 paper

The Mind Incarnate

Lawrence A. Shapiro

"Should lead scientists and philosophers to reconsider the pervasive assumption that the mind can be understood without reference to the material basis of mental processes."
— Elliott Sober, Stanford University

Life and Mind series • A Bradford Book
272 pp., 23 illus. $18 paper

Naturalistic Realism and the Antirealist Challenge

Drew Khlentzos

"Drew Khlentzos thinks that naturalistic realists have been far too complacent in the face of Dummett, Putnam, and others. In this wide-ranging and engaging book he tells us why. Very provocative!"
— Michael Devitt, The Graduate Center, City University of New York

Representation and Mind series
A Bradford Book • 448 pp., 10 illus. $20 paper

Consciousness and Persons

Unity and Identity

Michael Tye

"It would be impossible not to learn from this outstanding book."
— Frank Jackson, The Australian National University

Representation and Mind series
208 pp., 11 illus. $18 paper

True to Life

Why Truth Matters

Michael P. Lynch

"A bracing antidote to the disease of postmodern cynicism that renders truth impossible and leaves us with nothing but wind-blown opinion."
— *The Denver Post*

A Bradford Book • 224 pp. $14.95 paper

Natural Ethical Facts

Evolution, Connectionism, and Moral Cognition

William D. Casebeer

"Most defenders of evolutionary ethics will applaud [the book's] length, breadth, and eclecticism. . . . Highly recommended." — *Choice*

A Bradford Book • 224 pp., 2 illus. $19 paper

Three Philosophical Filmmakers

Hitchcock, Welles, Renoir

Irving Singer

"It is a lively pleasure to read Irving Singer's concrete and minutely attentive discussions of certain great films."
— Richard Wilbur, United States poet laureate, 1987-1988

312 pp., 4 illus. $16.95 paper

Philosophy from YALE

Plato's Republic

A Study

STANLEY ROSEN

"The most comprehensive and detailed commentary on and interpretation of Plato's *Republic* in English. Stanley Rosen is one of the most original interpreters of Plato alive."—Drew A. Hyland $45.00

The Philosophy of Positive Law

Foundations of Jurisprudence

JAMES BERNARD MURPHY

"An important and successful effort to trace the concept of positive law in both the natural law and legal positivist traditions, this book is an extraordinarily impressive integration of classical philosophy, political theory, legal history, and jurisprudence."—Frederick Schauer

$40.00

Kabbalah and Eros

MOSHE IDEL

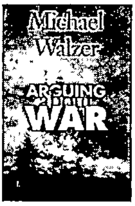
"Idel, with the brilliance and learning we have come to expect, has illuminated what went into 'the culture of eros' among a large cast of rabbinic and kabbalistic authors. His book makes for **surprising and fascinating reading."**—Geoffrey Hartman

$45 00

New in Paper

Arguing About War

MICHAEL WALZER



"Walzer has moved the concerns over just war from the periphery of political theory to the very center of our democratic dilemma."—Garry Wills, *New York Review of Books*

$16 00 paperback

The Enlightenment and the Intellectual Foundations of Modern Culture

LOUIS DUPRÉ

"Ranging as it does over art, morality, religion, science, philosophy, social theory, and a good deal besides, [this book] is **a marvel of scholarly erudition. . . .** A formidably well-researched book, which would make an excellent introduction to Enlightenment ideas for the general reader."

—Terry Eagleton, *Harper's Magazine*

$25 00 paperback

Leo Strauss and the Politics of American Empire

ANNE NORTON

"Like a backstage pass into the imperialistic mind of America, . . . [this book] trac[es] the evolution of Strauss's ideas as they seeped from the academy into the corridors of political power."—Mitchell Anderson, *San Francisco Bay Guardian*

$16.00 paperback

YALE University Press • yalebooks.com